UNIVERSITEIT GENT

FACULTY OF **SCIENCES**

Microbial diversity analyses in a changing landscape

Lactic acid bacteria in food fermentations as a test case

MSc. Isabel Snauwaert

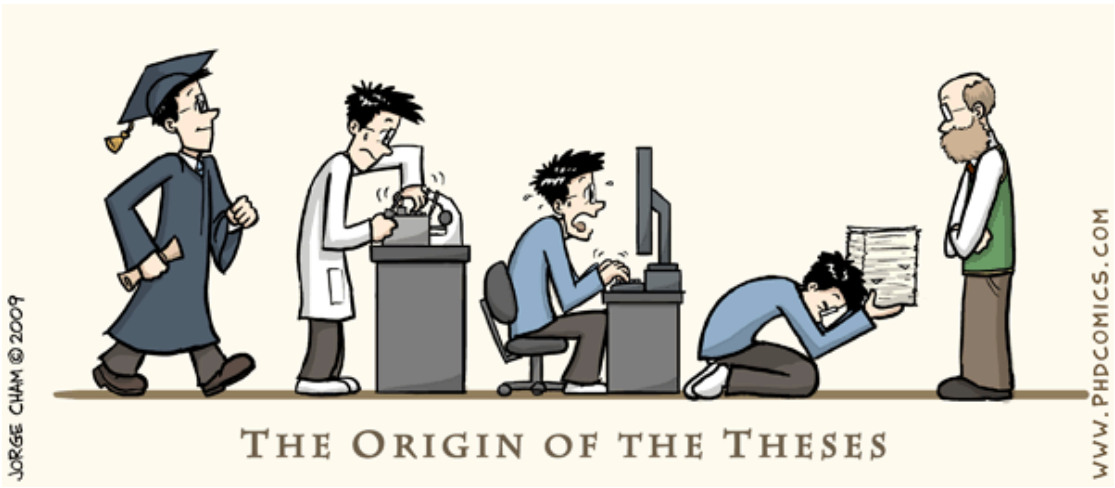
Dissertation submitted in fulfillment of the requirements for the degree of

Doctor (Ph.D.) in Sciences, Biochemistry and Biotechnology (Ghent University)

Doctor (Ph.D.) in Bioengineering Sciences (Vrije Universiteit Brussel)

Promotors: Prof. dr. Peter Vandamme (Ghent University)
Prof. dr. ir. Luc De Vuyst (Vrije Universiteit Brussel)
Prof. dr. Anita Van Landschoot (Ghent University)

I feel more microbe than man



Snauwaert, I. | Microbial diversity analyses in a changing landscape: Lactic acid bacteria in food fermentations as a test case

ISBN-number: 978-9-4619724-0-8

Printed by University Press | http://www.universitypress.be

Joint Ph.D. thesis, Faculty of Sciences, Ghent University, Ghent, Belgium
Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Brussels, Belgium

Publicly defended in Ghent, Belgium, November 25th, 2014

This Ph.D. work was supported by FWO-Flanders, BOF project, and the Vrije Universiteit Brussel (SRP, IRP, and IOF projects).

Author's email address: isabelsnauwaert@hotmail.com

Examination Committee

Prof. dr. Savvas SAVVIDES (Chairman)
L-Probe: Laboratory for Protein Biochemistry and Biomolecular Engineering
Faculty of Sciences, Ghent University, Ghent, Belgium

Prof. dr. Peter VANDAMME (Promotor UGent)
LM-UGent: Laboratory of Microbiology
Faculty of Sciences, Ghent University, Ghent, Belgium

Prof. dr. ir. Luc DE VUYST (Promotor VUB)
IMDO: Research Group of Industrial Microbiology and Food Biotechnology
Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel,
Brussels, Belgium

Prof. dr. Anita VAN LANDSCHOOT (Promotor UGent)
Laboratory of Biochemistry and Brewing
Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

Prof. dr. Edilbert VAN DRIESSCHE (Secretary)
SPRO: Research Group of Protein Chemistry
Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel,
Brussels, Belgium

Prof. dr. Tom COENYE
LPM: Laboratory of Pharmaceutical Microbiology
Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium

Prof. dr. Anne WILLEMS
LM-UGent: Laboratory of Microbiology
Faculty of Sciences, Ghent University, Ghent, Belgium

Prof. dr. ir. Ronnie WILLAERT
SBB: Research Group of Structural Biology Brussels
Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel,
Brussels, Belgium

Prof. dr. Tom LENAERTS
MLG: Machine Learning Group
Faculty of Sciences, Université Libre de Bruxelles, Brussels, Belgium

AI lab: Artificial Intelligence lab
Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel,
Brussels, Belgium

Prof. dr. Sebastiaan EELTINK
Research Group of Chemical Engineering
Faculty of Engineering, Vrije Universiteit Brussel, Brussels, Belgium

Prof. dr. Rudi VOGEL
Lehrstuhl für Technische Mikrobiologie
Center of Life and Food Sciences Weihenstephan, Technische Universität
München, München, Germany

Dankwoord

Vijf jaar heb ik als een 'Bezig Bijtje' in de gangen van het laboratorium voor Microbiologie rondgecrosst. Al die jaren heb ik mezelf kunnen zijn en me zo maximaal kunnen ontplooien. Mijn dankbaarheid uitdrukken op papier is dus geen eenvoudige opdracht en is sowieso ontoereikend. Maar ik doe hier toch een poging.

Professor Vandamme, Peter, allereerst wil ik u bedanken omdat u me de kans hebt gegeven om dit onderzoek uit te voeren. Bedankt om vertrouwen te hebben in mijn kunnen, om mij vrijheid te geven in de invulling van het project en om mij naar vele workshops en congressen te laten gaan. Dankzij u heb ik veel kunnen bijleren. *Professor De Vuyst*, bedankt om mij als één van uw doctoraatstudenten te behandelen. Zo nodigde u me uit om samen te lunchen met het IMDO-team en zelfs bij het kerstetentje, waardoor ik me bijzonder welkom voelde. *Professor Van Landschoot*, dankzij uw connecties met de brouwerijen konden we de nodige stalen verzamelen. Simpelweg onmisbaar voor mijn onderzoek.

Lieve buurvrouwtjes, jullie waren mijn warme thuisbasis binnen het labo. Onze bureau was een plaatsje waar ik veel gelachen heb, waar ik raad kon vragen, waar ik iedere morgen goed gezind werd en waar ik ook even triest mocht zijn. Jullie zijn stuk voor stuk top-collega's en vriendinnen. *Charlie*: mijn partner in crime én mijn tweede brain. Nooit gedacht dat iemand mijn gek idee ook effectief mee ging uitvoeren: op een workshop in Detroit om 4 uur in de ochtend opstaan om nog voor de start te kunnen fitnessen, ontbijten én scriptjes schrijven. En dan s' avonds vroeg in bed, want er was toch niets te doen in Detroit. Gelukkig wel in Liverpool: de cafeetjes waren er zo gezellig en het bier zeer drinkbaar. Maar nog belangrijker: brainstormen met jou was pure synergie, we brachten elkaar steeds tot een hoger niveau. Dankzij jou was er geen probleem dat niet tot een oplossing kon komen.

Anne, Timo, Jindrich, Freek, Jessy, Anneleen, Bram V., en *Helen*, mijn maatjes, veel meer dan alleen collega's. Samen gaan eten, op café gaan en kletsen kon alle dipjes doen overwaaien en gaf mij weer moed. Heerlijk om jullie te leren kennen! *Gwen, Sofie, Joke*, en *Jonas*: wat deed het pijn om jullie één voor één te zien vertrekken na het beëindigen van jullie doctoraat. *Gwen*, zonder jou was ik zelfs nooit doctorandus geworden. Je hebt als mijn stagebegeleister je enthousiasme, gedrevenheid en relativeringsvermogen op mij geprojecteerd. Door jou kreeg ik definitief de microbiologie-microbe te pakken. *Sofie*, jij was en bent nog steeds een onuitputtelijke bron van ideeën voor leuke activiteiten. En *Joke* en *Jonas*, ook jullie bleef ik tot op vandaag missen. *Kim*, bij de start van mijn doctoraat had ik het gevoel had dat je niet volledig geloofde in mijn capaciteiten. Maar achteraf ben ik daar blij om. Want zo heb je me assertiever en sterker gemaakt, want ik wilde bewijzen dat ik het wel kon. Een waardevolle attitude die ik nu meeneem naar het bedrijfsleven. En ook je lieve en dankbare tips voor de combinatie werk én zorg voor een pittige dochter, zullen mij altijd bij blijven.

Het schrijven van dit dankwoord toont aan dat er zoveel warme persoonlijkheden bij LM-Ugent werken. Zoals al eerder gezegd: dit dankwoord is sowieso ontoereikend. Zo denk ik nog aan *Jeannine, Evie, Bart H., Liesbeth, Maarten, Leilei, Bart V., en Joke* en alle andere collega's van LM-Ugent. Bedankt voor de warme en hartelijke sfeer in het labo, ik was er elke dag opnieuw welkom.

Lieve IMDO collega's, als bezoeker in een 'vreemd' laboratorium komen is niet evident. Ik was plots weer de leek die niets wist liggen en iedereen lastig viel met mijn vragen. Maar ik kon steeds bij jullie terecht. Expliciet een merci aan *Tom, Audrey, Maria, David, Henning, Moens* en *Koen*. Na twee uur treinen van Brugge naar Etterbeek voelde ik me vaak suf (zelfs koffie hielp dan niet). Maar na aankomst, kregen jullie mij altijd weer vrolijk.

Het doctoraat mocht dan wel boeiend zijn, maar gelukkig is er in mijn leven ook nog meer. Lieve *Lindseys*, het is altijd heerlijk om samen met mijn vriendinnen wijntjes te drinken en te kletsen. Jullie zijn stuk voor stuk topwijven. *Ilse*, mijn capaciteit om te plannen en efficiënt te werken was even zoek geraakt de laatste twee maanden. Merci om mij toen weer op de rails te zetten. En ook jij bedankt, *mevrouw Mitchell*, omdat je gewoon dé meest attente persoon bent die ik ken.

Mama en papa, bedankt voor alle kansen die jullie mij gaven. Jullie bleven mij onvoorwaardelijk steunen en geloofden in mijn capaciteiten. Door jullie kon ik studeren wat ik maar wilde en kan ik nu terugkijken op een heerlijke studententijd. Kortom, zonder jullie liefde en vertrouwen kan ik niet. *Papa*, ooit lachte ik met je als academische 'nerd' en 'pietje precies', maar ik vrees dat ik deze eigenschappen zeker van je meekreeg. *Mama*, ik hoop dat ik van jou het 'mama-zijn' heb geërfd en voor mijn dochter Helena nu net zo'n supermama zal zijn zoals jij altijd voor mij bent geweest (en nog altijd bent).

Ook *Evelien* (mijn ideale zus), *Filip* (de eeuwige enthousiasteling), *Sanne* (R en latex kenner: heerlijk om iemand met deze vaardigheden in de familie te hebben), *Mathilde* (altijd paraat als babysit) en mijn schoonfamilie: - *Jan* (English Expert), *Hilde* (de onmisbare SOS oma), de gekke *Leens*, *Bert* en *Pieter* - : dikke merci!

Het zijn vijf leerrijke jaren geweest, maar toch ben ik enorm opgelucht dat ik eindelijk een punt achter dit werk mag zetten. Het is nu tijd om mij te focussen op de belangrijke dingen in het leven en dat zijn jullie -*Bram* en kleine *Helena*-. *Bram*, jij hebt gevoeld hoe zwaar het afwerken van het doctoraat op mijn schouders woog na de geboorte van ons klein spook. Ik heb je de voorbije maanden meer dan ooit nodig gehad en jij bent er steeds voor mij geweest. Jouw onvoorwaardelijke liefde is het mooiste dat ik heb mogen tegengekomen in mijn leven. *Lieve dochter*, mijn klein pittig vrouwtje, ik kijk ernaar uit om je te zien opgroeien en je heel erg veel liefde te geven.

Gent, 17 november 2014
Isabel Snauwaert

Table of Contents

Examination Committee i

Dankwoord iii

Table of Contents vii

List of Figures xi

List of Tables xiii

Nederlandstalige Samenvatting xxiii

English Summary xxv

I Introduction 1

Background & Objectives 3

Outline 5

II Literature Overview 7

1 Microbial Diversity Analyses of Food Fermentations 9
1.1 Traditional Methods 10
1.1.1 Plate Culturing 11
1.1.2 Cultivation-Independent Molecular Techniques 12
1.2 Microbial Diversity Analyses in the HT Sequencing Era 14
1.2.1 Platforms 15
1.2.2 (Meta)genomics 17

2 **Importance of LAB in Food Fermentations** 23

3 **LAB Taxonomy** 27
3.1 Phylogenetics . . . 28
3.2 Polyphasic Taxonomy . . . 31
3.3 Genomic Taxonomy . . . 32

III Experimental Work 35

4 **Traditional Workflow in Microbial Diversity Analyses** 37
4.1 Characterization of Strains of *Weissella fabalis* sp. nov. and *Fructobacillus tropaeoli* from Spontaneous Cocoa Bean Fermentations . 39
4.1.1 Introduction . . . 40
4.1.2 Methods, Results, and Discussion . . . 40
4.1.3 Conclusion . . . 49
4.1.4 Acknowledgements . . . 51
4.1.5 Supplementary Material . . . 52
4.2 *Leuconostoc rapi* sp. nov., Isolated from Sous-Vide Cooked Rutabaga 55
4.2.1 Introduction . . . 55
4.2.2 Methods, Results, and Discussion . . . 56
4.2.3 Conclusion . . . 61
4.2.4 Acknowledgements . . . 62
4.2.5 Supplementary Material . . . 62
4.3 *Carnobacterium iners* sp. nov., a Psychrophilic, Lactic Acid-Producing Bacterium from the Littoral Zone of an Antarctic Pond. . . . 63
4.3.1 Introduction . . . 64
4.3.2 Methods, Results, and Discussion . . . 64
4.3.3 Conclusion . . . 71
4.3.4 Acknowledgements . . . 72
4.3.5 Supplementary Material . . . 72

5 **Microbial Diversity Analyses in the HT Sequencing Era** 73
5.1 Microbial Diversity and Metabolite Composition of Belgian Red-Brown Acidic Ales . . . 74
5.1.1 Introduction . . . 75
5.1.2 Materials and Methods . . . 76
5.1.3 Results . . . 82
5.1.4 Discussion . . . 95
5.1.5 Acknowledgments . . . 100
5.1.6 Supplementary Material . . . 100

5.1.7 Data Analysis Pipeline 108
5.2 Comparative Genome Analysis of *Pediococcus damnosus* LMG 28219, a Strain Well-Adapted to the Beer Environment 111
5.2.1 Introduction 112
5.2.2 Materials and Methods 113
5.2.3 Results and Discussion 115
5.2.4 Conclusions 129
5.2.5 Acknowledgments 129
5.2.6 Supplementary Material 129

IV General Discussion and Future Perspectives 143

6 Considerations for Microbial Diversity Analyses 147
6.1 Traditional Methods 147
6.2 Methods Involving HT Sequencing 148
6.2.1 Pre-Sequencing Considerations 148
6.2.2 Post-Sequencing Considerations 152
6.3 Importance of Cultivation 156
6.4 Integrative Ecosystems Biology 159

7 LAB Taxonomy in a HT Sequencing Era 161
7.1 Polyphasic Taxonomy 161
7.2 Genomic Taxonomy 163
7.3 Integrating Genomics into LAB Taxonomy 164

8 Future Perspectives 167

V Bibliography 171

VI Curriculum Vitae 199

List of Figures

1 Overview of the techniques used for microbial community analyses of food samples 10
2 HT sequencing platforms 16
3 PCR as a target enrichment strategy 21

4 Major pathways active in LAB 24
5 Consensus dendrogram reflecting the phylogenetic relationships of a selection of LAB genera of the order *Lactobacillales* within the class *Bacilli* 25

6 Schematic overview of the major analytical approaches to phylogenetic tree building. 29

7 ML tree based on 16S rRNA gene sequences showing the phylogenetic relationships of strains M75^{T} and M622 among the type strains of all species of the genera *Weissella* and *Fructobacillus* . . 41
8 ML tree based on concatenated *pheS*, *rpoA*, and *atpA* gene sequences showing the phylogenetic relationships of strains M56, M622, M710, M1190, and M1588 among other members of the genus *Fructobacillus* 45
9 ML tree based on *pheS* gene sequences showing the phylogenetic relationships of strain M75^{T} among other members of the genus *Weissella* 46
10 ML tree of the 16S rRNA gene sequence of strain LMG 27676^{T} with all type strains of the genus *Leuconostoc* 57
11 ML tree of the concatenated *pheS*, *rpoA*, and *atpA* gene sequences of the closest relatives of strain LMG 27676^{T} 59
12 ML tree based on 16S rRNA gene sequences showing the phylogenetic relationship of strains LMG 26642^{T} and LMG 26641 with type strains of species of the genus *Carnobacterium* 65

13 ML tree based on concatenated *pheS*, *rpoA*, and *atpA* sequences showing the phylogenetic relationship of strains LMG 26642^T and LMG 26641 with the type strains of species of the genus *Carnobacterium*. 67

14 1/D of the bacterial (A) and fungal (B) datasets of brew samples from breweries 1, 2, and 3 84
15 Relative abundances of the most abundant OTUs (> 0.01% of the reads) of the bacterial (A) and fungal (B) communities 86
16 PCA plot (A) and heatmap (B) of the metabolite concentrations of breweries 1, 2, and 3, determined in brew samples from the end of the maturation phase 89
17 PCA plot (A) and heatmap (B) of the metabolite concentrations of the bottled beers from breweries 1, 2, and 3 93
18 Visualization of the OrthoMCL output comparing the number of unique and/or shared orthologs of *P. claussenii* ATCC BAA-344^T, *P. pentosaceus* ATCC 25745, and *P. damnosus* LMG 28219 121
19 Comparison of plasmid pPECL-8 from *P. claussenii* ATCC BAA 344^T and contigs 70, 14, and 90 of *P. damnosus* LMG 28219 122
20 Folate biosynthesis pathway reconstruction 125
21 Overview of the functions enriched in the genomes of LAB of beer origin compared to those of LAB not originating from beer 128

22 Schematic outlining of the general approaches to culturing microbial diversity 158

23 Microbial diversity analyses in a HT sequencing era 168

List of Tables

1 Comparison of HT sequencing platforms 18

2 *Weissella* and *Fructobacillus* strains included in the phylogenetic study and in the MALDI-TOF MS analysis. 43
3 Differential characteristics of strain M75^T (*W. fabalis* sp. nov.) and other species of the genus *Weissella* 50
4 Differential phenotypic characteristics of strain LMG 27676^T and closest related species of the genus *Leuconostoc* 61
5 Amplification and sequencing primers used for the MLSA of *Carnobacterium* strains 67
6 Members of the genus *Carnobacterium* examined in this study . . . 68
7 Differential characteristics of strain LMG 26642^T and members of the genus *Carnobacterium* that assimilate a limited range of carbohydrates 70

8 Overview of the *P. damnosus* LMG 28219 draft genome. 117
9 Comparison of genome characteristics of *P. damnosus* LMG 28219 with *P. pentosaceus* ATCC 25745 and *P. claussenii* ATCC BAA-344^T 120

List of Acronyms

A

A	aerobic
AAB	acetic acid bacteria
AAI	average amino acid identity
ACA	American coolship ales
ADI	arginine deiminase pathway
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ANI	average nucleotide identity
ANOSIM	analysis of similarity
ARISA	automated ribosomal intergenic spacer analysis
ATCC	American type culture collection
ATP	adenosine triphosphate
atpA	RNA polymerase α-subunit

B

bcaT	branched chain amino acid aminotransferase
BLAST	basic local alignment tool
bp	base pair
BPW	buffered peptone water

C

C.	*Carnobacterium*
CBA	columbia blood agar
CCA	canonical correlation analysis
CCMM	Moroccan coordinated collections of microorganisms
CDS	coding sequence
CECT	colección Espanõla de cultivos tipo
CAMERA	community cyberinfrastructure for advanced marine microbial ecology research and analysis
CNS	coagulase-negative staphylococci
Cq	quantification cycle

D

1/D	inverse Simpson diversity index
DDBJ	DNA data bank of Japan
DDH	DNA-DNA hybridization
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
DSMZ	Deutsche sammlung von mikroorganismen und zellkulturen
dps	glucosyl transferase
DYPA	general yeast agar isolation medium
DYPAX	DYPA supplemented with 50 ppm cycloheximide

E

EC	enzyme commission
ELSD	evaporative light-scattering detector
ENA	European nucleotide archive
EPS	exopolysaccharide
ERDF	European regional development fund

F

F.	*Fructobacillus*
FAME	fatty acid methyl ester
FISH	fluorescence *in situ* hybridization
FWO	fund for scientific research-Flanders

G

GC	gas chromatography
GEBA	genomic encyclopaedia of bacteria and archaea
GOLD	genomes online database
groEL	molecular chaperone *groEL* gene
GSC	genomic standards consortium
gtf	transmembrane glycosyl transferase
GTP	guanoside triphospate
GTR	general time reversible

H

HMM	hidden markov model
HPAEC	high-performance anion exchange chromatography
HPLC	high-performance liquid chromatography
HT	high-throughput

I

IDT	integrated DNA technologies
IMG	integrated microbial genomes
IMG-ER	IMG expert review
IMG/M	IMG/metagenomes
INSDC	International nucleotide sequence databases collaboration

IS	insertion sequence
ITS	internal transcribed spacer

J

JCM	Japan collection of microorganisms

K

KCTC	Korean collection for type cultures
KEGG	kyoto encyclopedia of genes and genomes

L

L.	*Leuconostoc*
LAB	lactic acid bacteria
LaCOG	*Lactobacillales*-specific clusters of orthologous group
Lb.	*Lactobacillus*
LMG	laboratory of microbiology, Ghent University

M

mA	microaerobic
MALDI	matrix-assisted laser desorption/ionization
MEGA	molecular evolutionary genetics analysis
MEGAN4	metagenome analyzer
METAREP	meta gene annotator
MG	metagenome
MG-RAST	metagenomic rapid annotation using subsystems technology
MID	multiplex identifier
MIxS	minimum information about any sequence

ML	maximum-likelihood
MLSA	multilocus sequence analysis
MP	maximum parsimony
MRS	de Man-Rogosa-Sharpe
MS	mass spectrometry
MUM	maximal unique matches

N

NADP(H)	nicotinamide adenine dinucleotide phosphate
NCBI	National center for biotechnology information
NCFB	National collection of food bacteria
NCIMB	National collections of industrial, food and marine bacteria
ND	no data available
NK	not known
NJ	neighbour joining
no.	number

O

OTU	operational taxonomic unit

P

P.	*Pediococcus*
pABA	4-aminobenzoic acid
PATRIC	pathosystems resource integration center
PCA	principal component analysis
pcDNA	percentage of conserved DNA
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
PE	paired-end
pepN	aminopeptidase N
pepX	X-prolyl dipeptidyl peptidase

PGAAP	prokaryotic genome automatic annotation pipeline
PHAST	phage search tool
pheS	phenylalanyl-tRNA synthase α subunit
P(P)i	(pyro)phosphate

Q

QIIME	quantitative insights into microbial ecology
qPCR	quantitative PCR

R

RAPD	randomly amplified polymorphic DNA
RAST	rapid annotation using subsystems technology
RDP	ribosomal database project
rep-PCR	repetitive element sequence-based PCR
RNA	ribonucleic acid
rpoA	RNA polymerase α subunit
rpoB	β-subunit of the bacterial RNA polymerase
rRNA	ribosomal RNA

S

SCFA	short-chain fatty acids
SE	standard error
SH	static-headspace
sp.	species
SRA	sequence read archive

T

T	type strain

TLC	thin layer chromatography
TOF	time-of-flight
TRF	terminal restriction fragment
TRFLP	terminal restriction fragment length polymorphism
tRNA	transfer RNA
TSA	tryptone soya agar
TSBY	trypticase soy broth with yeast extract salt medium

U

UPGMA	unweighted pair group method with arithmetic means
UV	ultra violet

V

VBNC	viable but non-cultivable

W

W.	*Weissella*

Nederlandstalige Samenvatting

Micro-organismen, inclusief melkzuurbacteriën, dragen bij tot de smaak, het aroma, de textuur, en de houdbaarheid van gefermenteerde levensmiddelen en dranken, die deel uitmaken van het dagelijks dieet van de mens. Kennis over de microbiële gemeenschappen in deze fermenterende ecosystemen is nodig om veilige producten met gewenste eigenschappen te kunnen afleveren aan de consument. Het doel van deze thesis is het evalueren van de huidige methoden om de microbiële diversiteit van voedsel ecosystemen te analyseren en om nieuwe soorten van melkzuurbacteriën te beschrijven.

Een grote fractie aan micro-organismen kunnen niet gekweekt worden door middel van conventionele cultivatie-condities, wat leidt tot een vertekend beeld van de werkelijke microbiële diversiteit van een ecosysteem wanneer cultivatie-afhankelijke methoden worden aangewend. Cultivatie-onafhankelijke moleculaire technieken (zoals '*denaturing gradient gel electrophoresis*', '*terminal restriction fragment length polymorphism*', '*clone libraries*', enz.) zijn niet afhankelijk van de kweekbaarheid van micro-organismen in een gemeenschap, maar deze zijn ongeschikt door hun arbeidsintensiviteit, tijdrovende natuur, en lage-throughput. Om deze redenen zijn effectieve methoden essentieel die het documenteren van de microbiota in levensmiddelen mogelijk maken.

De ontwikkeling van high-throughput sequenering methoden en de resulterende sequenering aan lage kost per base hebben het microbiologisch onderzoek de laatste decennia drastisch veranderd. In deze thesis werd high-throughput sequenering toegepast om het ecosysteem van Belgische roodbruine bieren te bestuderen en om het genoom van het dominante bacteriële lid van deze gemeenschap te sequeneren. Daaruit blijkt dat high-throughput sequenering benaderingen waardevol zijn voor bestuderen van de diversiteit en het metabolische potentieel van micro-organismen. Toch, door fouten, geïntroduceerd tijdens microbiële diversiteit-studies die high-throughput sequenering toepassen, is het essentieel om deze te overwegen vooraleer een experiment aan te vatten en tijdens de analyse van de data.

De traditioneel toegepaste polyfasische taxonomische benadering werd in deze thesis aangewend om nieuwe soorten van melkzuurbacteriën te beschrijven. Daaruit blijkt dat deze benadering gehinderd is door haar arbeidsintensieve en tijdrovende natuur, wat contrasteert met de grote hoeveelheid stammen die momenteel wachten om beschreven te worden. Daarbovenop contrasteert de polyfasische taxonomische benadering met de opkomende '*culturomics*' aanpak, die het potentieel heeft om een grote hoeveelheid tot op heden niet gekarakteriseerde stammen te genereren. Om bovenstaande redenen is de polyfasische taxonomische benadering verouderd en moet deze dringend herbekeken worden om een snelle beschrijving van nieuwe soorten van melkzuurbacteriën in de toekomst mogelijk te maken.

Omdat volledige genoom sequenties inzichten verwerven in de genetische natuur van microbiële species, kunnen ze gebruikt worden om bacteriële species af te lijnen en om hun fylogenie te bestuderen. De beschrijving van nieuwe soorten in het genomische tijdperk zou een genoom sequentie en een minimale beschrijving van fenotypische eigenschappen moeten bevatten, die samen kunnen fungeren als een biologische identiteitskaart die voldoende, kostenbesparend, en geschikt is voor de beschrijving van nieuwe species. De opslag en het onderhoud van deze biologische identiteitskaarten zal resulteren in taxonomische databanken van hoge kwaliteit die de wetenschappelijke wereld van dienst kunnen zijn om de kwaliteit van hun annotaties en taxonomische toewijzingen te verbeteren. Bovendien zouden bacteriële taxonomen kunnen bijdragen aan de inspanningen van het '*Genomic Encyclopaedia of Bacteria and Archaea*'-project om de afwijkende fylogenetische verdeling van de beschikbare genoom-sequenties in te perken.

Finaal zou het combineren van (meta)genoom-, (meta)transcriptoom-, (meta)proteoom-, en (meta)metaboloom-data met cultivatie nieuwe inzichten kunnen verwerven in de interacties tussen de leden van een microbiële gemeenschap. Bovendien zou deze geïntegreerde aanpak belangrijk kunnen zijn om interventies te ontwerpen die gefocust zijn op de functies van de gemeenschappen in levensmiddelen in plaats van de afzonderlijke species.

English Summary

Microorganisms, including lactic acid bacteria, contribute to the taste, aroma, texture, and shelf life of fermented foods and beverages, which are part of the human diet globally. Knowledge on the microbial communities within these fermenting ecosystems is required to deliver safe products with desirable properties. The goals of this thesis were to evaluate the state-of-the-art approaches for microbial diversity analyses of food ecosystems and for the description of novel lactic acid bacteria taxa.

Because a large fraction of the microbial consortia cannot be cultured using conventional growth conditions, culture-dependent approaches lead to a biased view on the actual microbial diversity. Culture-independent molecular techniques (*i.e.*, denaturing gradient gel electrophoresis, terminal restriction fragment length polymorphism, clone libraries, and others) do not depend on the cultivability of microorganisms in a community, but they are inadequate for large-scale microbial diversity analyses because of their labor-intensity, low-throughput, and time-consuming nature. Therefore, effective tools are required enabling thorough documentation of the microorganisms in food ecosystems.

The advent of high-throughput sequencing technologies and the resulting sequencing at reduced cost per base have changed microbiological research drastically in the last decade, making high-throughput sequencing affordable for many research laboratories. In this thesis, high-throughput sequencing technologies were applied to unravel the microbial diversity of mature Belgian red-brown acidic ales and to sequence the genome of the dominant bacterial member of this beer ecosystem. These high-throughput sequencing technologies proved to be valuable for studying the diversity and the metabolic potential of environmental microorganisms. Nevertheless, because of the biases associated with high-throughput sequencing, it is crucial to consider these before launching an experiment and while handling the data.

The traditionally used polyphasic taxonomy approach was applied to describe novel lactic acid bacteria species and proved to be labor-intensive and thus unsuited to perform in high-throughput, which is contradictory with the huge amount of strains that are currently awaiting description and formal naming. Furthermore, this approach contrasts with the emerging culturomics method, which has the potential to generate a vast amount of not yet characterized strains. As a consequence, the polyphasic taxonomy approach is outdated and should urgently be reconsidered to enable fast description of novel lactic acid bacteria taxa.

Because whole-genome sequences provide insights into the genetic nature of microbial species, they can be used as a tool for delineating bacterial species and for studying their phylogeny. The genomic taxonomy approach should include a whole-genome sequence and a minimal description of phenotypic characteristics, which will serve as a basic biological identity card that be considered sufficient, cost-effective, and appropriate for species descriptions. These biological identity cards should be stored and maintained in high-quality, readily accessible, iterative, and adaptable taxonomic databases, which can serve the scientific community to improve the quality of their annotations and taxonomic assignments. Furthermore, bacterial taxonomists could contribute to the phylogenetically driven sequencing effort, initiated by the Genomic Encyclopaedia of Bacteria and Archaea project, and thereby reduce the biased phylogenetic distribution of the currently available genome sequences.

Finally, combining (meta)genome, (meta)transcriptome, (meta)proteome, and (meta)-metabolome approaches with cultivation could provide new insights into relations between members of a microbial community and will be important for designing interventions targeted at functions of the microbial community in food rather than specific constituent species.

Part I

Introduction

Background & Objectives

Microorganisms, including lactic acid bacteria (LAB), contribute to the taste, aroma, texture, and shelf life of fermented foods and beverages, which are part of the human diet globally. Knowledge on the microbial communities within these fermenting ecosystems is required to deliver safe products with desirable properties. A large fraction of these microbial consortia cannot be cultured using conventional cultivation conditions, leading to a biased view on the actual microbial diversity when applying traditional approaches. Therefore, effective tools are required enabling documentation of the microorganisms in food ecosystems. Developments in high-throughput (HT) sequencing technologies provide a world of opportunities for investigating these ecosystems. Furthermore, the currently used polyphasic taxonomy approach is hampered by its labor-intensive and time-consuming nature, and microbiologists are therefore looking for alternatives based on the evolutionary information contained in whole-genome sequences.

The first goals of the present study were to evaluate **(i)** the state-of-the-art approaches for microbial diversity analyses applied to food ecosystems and **(ii)** the polyphasic taxonomy approach currently used for the description of novel species. Furthermore, **(iii)** the existing multilocus sequence analysis (MLSA) approach was extended to the genus *Carnobacterium*, which includes food-grade LAB species. Next, **(iv)** a HT sequencing approach, comprising targeted sequencing of the bacterial partial 16S ribosomal ribonucleic acid (rRNA) gene and the fungal internal transcribed spacer (ITS) region, was applied to study the ecosystem of mature Belgian red-brown acidic ales. The goal was to update and broaden the knowledge on a previously studied Belgian red acidic ale and to compare its microbial community diversity and metabolite composition with two additional Belgian red-brown acidic ales. Following, **(v)** whole-genome sequencing of *Pediococcus damnosus* LMG 28219, which was the dominant LAB member of the Belgian red-brown acidic ale ecosystem, and comparative genome analysis were used to investigate this strain's mechanisms to reside in the beer niche. Finally, **(vi)** the value of HT sequencing technologies in microbial diversity analyses was evaluated and the advantages of incorporating genome information into the field of taxonomy were explored.

Outline

In **PART II** of the present thesis, a comprehensive LITERATURE OVERVIEW is given on microbial diversity analyses in a changing landscape, with special focus on LAB in food fermentations. **Chapter 1** describes the developments of molecular technologies for investigating the diversity of microorganisms in fermented foods and beverages, with a particular emphasis on the application of novel HT sequencing methods. **Chapters 2 and 3** review the importance of LAB in food fermentations and their taxonomy. **Chapter 3.1** provides background information on phylogenetics.

PART III presents the EXPERIMENTAL WORK performed during this thesis. In **Chapter 4**, the traditional workflow for microbial diversity analyses was applied to the spontaneous cocoa bean fermentation ecosystem and the sous-vide cooked rutabaga ecosystem. Novel LAB species (*i.e.*, *Weissella fabalis* sp. nov and *Leuconostoc rapi* sp. nov) discovered during these analyses, were described using a polyphasic taxonomy approach. Furthermore, the MLSA scheme for the genus *Carnobacterium* was completed, along with a polyphasic description of *Carnobacterium iners* sp. nov. **Chapter 5** comprises the application of HT sequencing technologies to unravel the microbial diversity of Belgian red-brown acidic ales, along with a comparative genome analysis of the *Pediococcus damnosus* LMG 28219 dominant LAB species, isolated from a Belgian red acidic ale, with other *Pediococcus* genome sequences.

PART IV presents a GENERAL DISCUSSION of the results and provides FUTURE PERSPECTIVES.

Part II

Literature Overview

1

Microbial Diversity Analyses of Food Fermentations

Fermentation is a biotechnological method historically arisen from the need to preserve food. It can be carried out by bacteria, yeasts, and/or filamentous fungi, which convert fermentable carbohydrates mainly into organic acids, alcohols, and/or carbon dioxide (Leroy & De Vuyst, 2004; Ravyts *et al.*, 2012). LAB play a central role in the production of fermented foods and beverages, because of their production of lactic acid as a common end-metabolite (Stiles & Holzapfel, 1997). Next to LAB, other bacterial species have importance in food fermentations. For instance, acetic acid bacteria (AAB) contribute to the production of vinegar and cocoa (Raspor & Goranovic, 2008). Furthermore, coagulase-negative staphylococci (CNS), *Kocuria*, and *Micrococcus* species are indispensable natural microbiota of fermented meat products next to lactobacilli and pediococci (Leroy *et al.*, 2006) and can also be found in both hard and soft cheeses (Irlinger, 2008). Other bacterial groups with importance in food fermentations include brevibacteria, corynebacteria, and propionibacteria, which are used in cheese production, and *Bacillus subtilis* that is used for soybean fermentations (Ravyts *et al.*, 2012). In carbohydrate-rich environments, acidifying LAB frequently grow in association with yeasts that are present in lower numbers. Yeast species of *Saccharomyces*, *Candida*, *Torula*, *Hanseniaspora*, *Hansenula*, *Dekkera*, and others, can proliferate in these niches, causing spontaneous alcoholic fermentation of wine and beer (Fleet, 2007). Finally, filamentous fungi, such as *Aspergillus* and *Rhizopus* may play a role, such as in cheeses, fermented sausages, and soy-based fermentations (Cocolin *et al.*, 2013).

Searching for the presence, numbers, and type of microorganisms in foods is of great importance for the food industry. Whatever the primary objective of these microbial analyses (*e.g.*, control of food quality or preservation, efficiency of starter cultures, monitoring of particular species/strains, detection of unwanted microorganisms), the taxonomic level of the microbial discrimination needed should be initially decided. An overview of the methods used in microbial diversity analyses is presented in Figure 1; the traditionally applied methods are discussed in Section 1.1, whereas those involving HT sequencing are elaborated in Section 1.2.

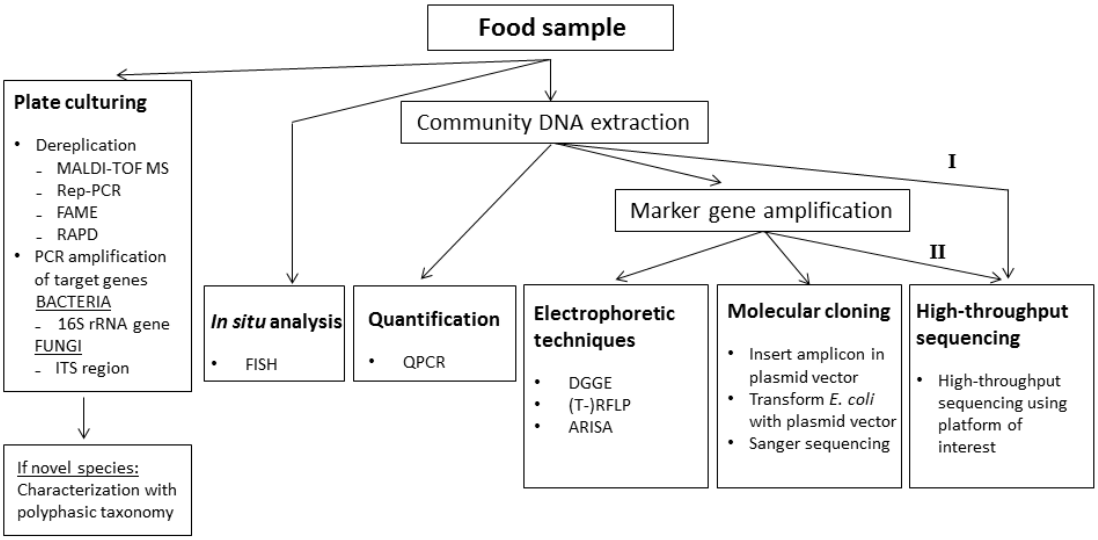


Figure 1: Overview of the techniques used for microbial community analyses of food samples

I: Shotgun sequencing strategy, II: Target enrichment strategy. Abbreviations are explained in Section 1.1.

1.1 Traditional Methods

A variety of culture-dependent and culture-independent techniques are traditionally used in microbial diversity analyses to study the phylogenetic and functional diversity of food ecosystems.

1.1.1 Plate Culturing

Microbial diversity analyses traditionally involve the cultivation of microorganisms onto a set of general and selective growth media and conditions. Following incubation, a dedicated percentage of colonies with morphologically distinct types are randomly picked up -whether or not statistically relevant- for further analysis. As bacteria are a convenient source of intrinsic marker proteins, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is an appropriate tool for the HT generation of strain-specific spectral profiles (Freiwald & Sauer, 2009; Ghyselinck *et al.*, 2011). By comparing these spectral profiles, a large set of isolates can easily be grouped, which allows the selection of a smaller set of representative strains, the so-called dereplication step (Ghyselinck *et al.*, 2011). Dereplication is important in large-scale isolation campaigns and screening programs, since it can significantly reduce labor intensity and time, and costs in further downstream analyses. Other tools used in the context of dereplication are repetitive element sequence-based polymerase chain reaction (rep-PCR) (De Clerck & De Vos, 2002), randomly amplified polymorphic deoxyribonucleic acid (RAPD) (Martín-Platero *et al.*, 2009), and fatty acid methyl ester (FAME) (Coorevits *et al.*, 2008) analyses. The resulting reduced set of isolates is assumed to be representative for the microbial communities studied and is subsequently identified using Sanger sequencing (Sanger *et al.*, 1992) of marker genes. Alternatively, a tentative identification can be obtained by querying MALDI-TOF MS protein profiles of this reduced set against an in-house developed or a commercial reference database containing high-quality spectra (Andres-Barrao *et al.*, 2013; Clark *et al.*, 2013; De Bruyne *et al.*, 2011; Dieckmann *et al.*, 2005; Duskova *et al.*, 2012; Wieme *et al.*, 2014). Examples of microbial diversity analyses of food ecosystems involving cultivation are numerous [as example (Spitaels *et al.*, 2014b; Papalexandratou *et al.*, 2011a; Laureys & De Vuyst, 2014; Pothakos *et al.*, 2014; Nguyen *et al.*, 2013)].

This labor-intensive plate culturing approach gives a biased view on the actual microbial diversity, because of the inability of certain community members to grow on the media or in the incubation conditions selected, and because of the challenging recovery of microorganisms, which are either stressed or entered into a viable but non-cultivable (VBNC) state (Fleet, 1999; Rappé & Giovannoni, 2003). The VBNC state is induced when adverse conditions such as nutrient depletion, low temperatures, and stresses cause cultivable cells to enter a phase, in which they are still capable of metabolic activity but do not produce colonies on media that normally would support their growth. As a consequence, the cultivation-based approach highlighted above is inadequate for the complete characterization of microbial communities.

1.1.2 Cultivation-Independent Molecular Techniques

Many of the cultivation-independent molecular methods for microbial diversity studies rely on the analysis of the 16S rRNA gene. This gene is part of the ribosomal operon and its size, nucleotide sequence, and secondary structure is conserved within bacterial species (Maidak *et al.*, 1997). It is composed of highly conserved sequence domains interspersed with hypervariable regions, which serve as primer docking sites and as phylogenetic markers, respectively. However, the presence of multiple rRNA operons in the bacterial genome of some species may lead to an overestimation of the microbial richness (Acinas *et al.*, 2004; Klappenbach *et al.*, 2000). Furthermore, the 16S rRNA gene does not reflect the remaining functional genome content (Robinson *et al.*, 2010). These restrictions have led to the consideration of alternative single-copy housekeeping genes to study the bacterial diversity of microbial ecosystems, such as the *rpoB* gene, encoding the β-subunit of the bacterial RNA polymerase (Case *et al.*, 2007). However, no gene is without limitations and the lack of universal primers and comprehensive databases for a wide spectrum of strains hampers their implementation in microbial community surveys. Similar to the bacterial 16S rRNA gene, the preferred molecular target for fungal phylogenetics is the ITS region of the nuclear DNA (Seifert, 2009).

Popular community profiling methods for food ecosystems are denaturing gradient gel electrophoresis (DGGE) and terminal restriction fragment length polymorphism (TRFLP) of amplified fragments. Although only sporadically used by food microbiologists, automated ribosomal intergenic spacer analysis (ARISA) can also be applied for community profiling. Those methods only provide a qualitative (*i.e.*, presence or absence) assessment of microbial communities. DGGE, TRFLP, and ARISA consist of the PCR amplification of a marker gene fragment of interest from the community DNA, followed by separation of the amplicons. Therefore, these methods are prone to amplification-related biases common to all PCR-based techniques. In the case of DGGE, the separation of PCR amplicons is based on a decreased electrophoretic mobility of PCR-amplified, partially melted, double-stranded DNA molecules in polyacrylamide gels containing a linear gradient of a DNA denaturant. Molecules with different sequences may have a different melting behavior and will stop migrating at different positions along the gel. The patterns generated could provide a preliminary ecological view of the predominant community members (Muyzer *et al.*, 1993). DGGE is applied by many researchers in the food microbiology field [as example (Gkatzionis *et al.*, 2014; Pangallo *et al.*, 2014; Papalexandratou *et al.*, 2011a; Papalexandratou & De Vuyst, 2011b; Papalexandratou *et al.*, 2011d)]. The drawbacks associated with this method are that it is time-consuming, has a low-throughput capacity, and is technically challenging, as the denaturing gradient gels are difficult to set and run properly (Papalexandratou *et al.*, 2011a). In addition, prokaryotic 16S rRNA genes may contain insufficient heterogeneity to reliably identify some bacterial species using the short amplicons necessary for DGGE separation. Another popular community profiling method to study microbial ecology

is TRFLP, which has received less attention by food microbiologists. This method differentiates microbial community members based on terminal restriction fragment (TRF) sizes of the amplified regions. Mixed DNA templates are amplified with fluorescence-labeled universal primers, digested with selected restriction enzymes, and separated by capillary electrophoresis according to fragment size. This method has been applied to study cheese and yogurt fermentation (Rademaker *et al.*, 2006), to analyze viable bacterial communities during cheese fermentations (Sanchez *et al.*, 2006), for species-level LAB differentiation (Bokulich & Mills, 2012c), to profile microbial successions during beer and wine fermentations (Bokulich *et al.*, 2012a), to analyze the bacterial composition of commercial probiotic strains (Marcobal *et al.*, 2008), and others. This method is HT and has a low technical demand, but it encounters the same problems as DGGE regarding the low taxonomic depth of the identifications obtained. Finally, ARISA is based on the use of a fluorescent primer for the amplification of microbial ribosomal intergenic spacers. The PCR products are analyzed in an automated capillary electrophoresis system that produces an electropherogram, the peaks of which correspond with discrete DNA fragments detected by a laser-based fluorescence detection system. The sensitivity of ARISA is very high and its reproducibility is reinforced by instrumental automation. This method is not widely used in food microbiology because the primer sets described are not optimized for food samples and the lack of large datasets comprising intergenic spacer sequences, limiting the identification of these fragments. It has been applied by Panelli and coworkers (2013) to diagnose the presence of *Clostridium tyrobutyricum* in raw milk before cheese making.

Although more expensive and time-consuming than the community fingerprinting techniques highlighted above, sequence analysis of clone libraries provides an unparalleled level of phylogenetic resolution due to the long read lengths generated by Sanger sequencing technology (Sanger *et al.*, 1992). This method involves PCR amplification of phylogenetic marker genes, followed by insertion of these PCR products into a plasmid vector, which is subsequently used for transformation of competent *Escherichia coli* cells. The DNA sequence of the insert is finally resolved using the Sanger sequencing technology. Clone libraries have been constructed in numerous studies on food fermentation, including the detection of cucumber fermentation spoilage bacteria (Breidt *et al.*, 2013), the assessment of the bacterial community structure in kimchi (a Korean fermented vegetable food) (Kim & Chun, 2005), sour cassava starch, cachaça, and minas cheese production (Lacerda *et al.*, 2011), Ghanaian and Brazilian cocoa bean fermentations (Garcia-Armisen *et al.*, 2010), and others. The drawbacks associated with clone libraries are that the method is time-consuming, has a low-throughput capacity, is labor-intensive, and is prone to PCR biases. Furthermore, it can only provide a qualitative assessment of the microbial communities.

The quantification deficit of the approaches discussed above can be counteracted by quantitative PCR (qPCR) analysis, a powerful extension of conventional PCR that allows reliable and accurate quantifications. qPCR has become an extremely

popular method in food microbial ecology studies, in which amplification is tracked in real-time using a fluorescent reporter molecule. A baseline threshold is determined at which sample fluorescence can be distinguished from background noise. This is used to determine the quantification cycle (C_q) for each sample, at which fluorescence crosses the baseline threshold. A standard curve is constructed by plotting C_q against gene copy number or cell count and is used to quantify unknown samples based on C_q. Beyond cell quantification based on a taxonomic marker gene, the detection of functional genes provides interesting applications in food microbiology as well. This approach has been applied to detect biogenic amine-producing strains of LAB in wines (Arena *et al.*, 2011; Lucas *et al.*, 2008), ciders (Ladero *et al.*, 2011), and cheeses (Fernandez *et al.*, 2006; Torriani *et al.*, 2008), to quantify exopolysaccharide (EPS)-producing LAB (Ibarburu *et al.*, 2010), to track sulfide reductase genes in *Saccharomyces cerevisiae* that are involved in off-flavor production (Mendes-Ferreira *et al.*, 2010), and others.

Another frequently applied technique to detect and quantify bacteria in food samples is fluorescence *in situ* hybridization (FISH), which is used for directly visualizing microbial cells in a sample. This method utilizes fluorescence-labeled oligonucleotide probes targeting specific DNA sequences, unique for discrete microbial taxa. Cells are fixed and permeabilized *in situ*, incubated in the presence of probes to facilitate hybridization, and subsequently observed directly by fluorescence microscopy or counted via flow cytometry if suspended in a fluid. This technique has become very popular, as it avoids the challenges and biases of culturing and PCR assays and, additionally, it allows the observation of target cells within their native environment. The drawbacks of this method are its low-throughput capacity and its lower efficiency for enumerating cells compared to qPCR. FISH has been used to monitor microbial communities in a number of fermented foods, including yeasts and bacteria in wine (Andorra *et al.*, 2011; Blasco *et al.*, 2003; Xufre *et al.*, 2006), beer (Yasuhara *et al.*, 2001), and cheese (Babot *et al.*, 2011; Bottari *et al.*, 2010; Mounier *et al.*, 2009).

1.2 Microbial Diversity Analyses in the HT Sequencing Era

The traditional Sanger DNA-sequencing technology (Sanger *et al.*, 1992) can only sequence specimens individually and is therefore inadequate for processing complex microbial communities, especially in large-scale studies. Although conventional sequencing has provided large DNA reference databases, the number of individuals in a complex sample is beyond its ability. Recovering DNA sequences from complex and/or multiple samples requires the capacity to read DNA from multiple templates in parallel, which HT sequencing technologies do effectively at low costs.

1.2.1 Platforms

In 2005, the first HT sequencing technology emerged and drastically reduced the time and cost needed for sequencing (Metzker, 2005). The currently available HT sequencing technologies can be categorized into two broad groups, depending on the kind of template needed for the sequencing reactions (Loman *et al.*, 2012). The oldest group comprises the construction of a library of clonally amplified templates [*e.g.*, the 454 (Roche Diagnostics), Ion Torrent (Life Technologies), Illumina (Illumina Inc.), SOLiD (Life Technologies), and Complete Genomics (Complete Genomics Inc.) platforms], whereas the more recently developed group of HT sequencing technologies involves the sequencing of single molecules [*e.g.*, the HeliScope Single-Molecule (Helicos BioSciences), PacBio RS (Pacific Biosystems), and Oxford Nanopore (Oxford Nanopore Technologies) platforms] (Figure 2). Considerable variation in cost, time, throughput, read length, and error rate exists within these broad categories of HT sequencing technologies.

The first group of template amplification technologies involves a three-stage workflow of (i) library preparation, (ii) template amplification, and (iii) sequencing (Figure 2). Library preparation (i) for shotgun sequencing involves an initial (mechanical or enzymatic) fragmentation step to generate random, overlapping DNA fragments. Fragmentation is followed by the ligation of adapters to the fragments generated. Tagmentation is a promising transposase-based approach that, in a single step, fragments DNA and incorporates sequence tags. Currently, the only available implementation of tagmentation is within the Nextera® system, which is only available for the Illumina platform. Next to shotgun sequencing, all platforms can handle PCR products, allowing adapter sequences to be incorporated into the 5′ ends of the primers. In preparation of amplification, template molecules are immobilized on a solid surface, which is a flow cell for sequencing with the Illumina platform or beads for other approaches (ii). Simultaneous solid-phase amplification of millions or billions of spatially separated fragments, prepares the way for massive parallel sequencing. For the Illumina platform, clusters are generated by bridge amplification on the surface of a flow cell. For platforms using bead-based immobilization (the SOLiD, 454, and Ion Torrent platforms), amplified template sequence libraries rely on an emulsion PCR, in which the beads are enclosed in aqueous-phase microreactors and are kept separately from each other into a water-oil emulsion.

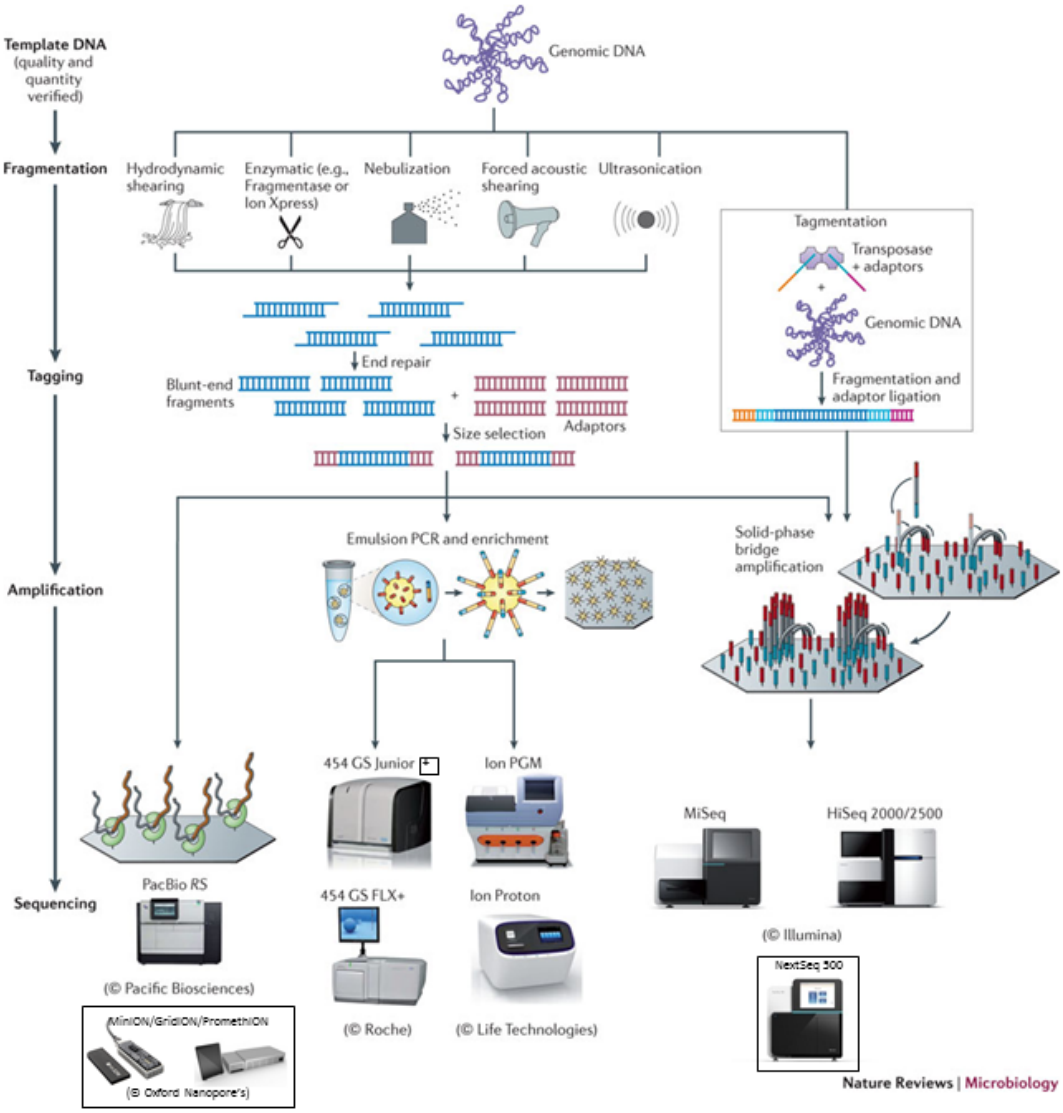


Figure 2: HT sequencing platforms

The main HT sequencing platforms available and their associated sample preparation and template amplification procedures are shown. PGM: personal genome machine.
Adapted from (Loman *et al.*, 2012). Adaptations to the original figure are highlighted with black squares.

The last part of the workflow in template amplification technologies is the sequencing (iii). Illumina, 454, and Ion Torrent technologies rely on the sequencing-by-synthesis design, whereas the SOLiD platform and the platform from Complete Genomics depend on sequencing-by-ligation. Illumina uses the Solexa chemistry that includes reversible termination of sequencing products, whereas in 454 and Ion Torrent a dNTP flows across the template in each cycle. More specifically, 454 uses the pyrosequencing approach, whereby the presence of pyrophosphate is signaled by visible light as the result of an enzyme cascade. In contrast, Ion Torrent does not rely on light emission, but uses a modified silicon chip to detect hydrogen ions that are released during base incorporation. The SOLiD and Complete Genomics platforms use an approach in which fluorescent probes undergo iterative steps of hybridization and ligation to complementary positions in the template strand followed by fluorescence imaging to identify the ligated probe.

The second group of single-molecule sequencing platforms does not comprise amplification and library preparation steps, thereby circumventing the artifact associated with these steps. The HeliScope Single-Molecule sequencer was the first entering the market in 2009. This technology applies one-color reversible terminator sequencing to unamplified single-molecule templates. Another more recent technology is the one of PacBio, in which dye-labeled nucleotides are continuously incorporated into a growing DNA strand by a specialized strand-displacing DNA polymerase that is attached to a solid surface. This technology allows continuous imaging of the labeled nucleotides as they enter the strand. Another promising single-molecule sequencing technology is the one of Oxford Nanopore, which is currently developing the 'strand sequencing' technology that exploits protein nanopores embedded in an industrially fabricated polymer membrane. This technology passes intact DNA polymers through a protein nanopore, allowing sequencing in real-time as the DNA translocates the pore.

Most companies have both high-end and bench-top instruments to comply with the needs of different users, which are large sequencing centers and research laboratories, respectively. The high-end machines include PacBio RS, the HiSeq instruments, Genome Analyser IIx, the SOLiD 5500 series, and the 454 GS FLX$^+$ system, whereas the bench-top instruments include the GS Junior, Ion PGM, and the Illumina MiSeq. An overview of these instruments and their performance is summarized in Table 1.

1.2.2 (Meta)genomics

The advent of HT sequencing technologies and the resulting sequencing at reduced cost per base have changed microbiological research drastically in the last decade, making genomics and metagenomics affordable for many research laboratories.

Table 1: Comparison of HT sequencing platforms

Machine (manufacturer)	Chemistry	Modal read length (bases)	Run time	Output per run	Advantages	Disadvantages
High-end instruments						
454 GS FLX+ (Roche)	Pyrosequencing	700 bp/450 bp	23 hours/10 hours	700 Mb/450 Mb	Long read lengths	Appreciable hands-on time, high reagent costs, high error rate in homopolymers
HiSeq2500 (Illumina)	Reversible terminator	2x250/2x125 bp	7-60 hours/<1 day - 6 days	10-300 Gb/50-1000 Gb	Cost-effectiveness, steadily improving read lengths, massive throughput, minimal hands-on time	Long run time, short read lengths
NextSeq500 (Illumina)	Reversible terminator	2x150 bp	15-26 hours/12-30 hours	20-39 Gb/30-120 Gb	Cost-effectiveness, steadily improving read lengths, massive throughput, minimal hands-on time	Long run time, short read lengths
PacBio RS (Pacific Biosystems)	Real-time sequencing	>14000 bp	30 -240 minutes	500 Mb-1 Gb	Simple sample preparation, low reagent costs, very long read lengths	high error rate, expensive system, difficult installation
MinION, PromethION, GridION (Oxford Nanopore's)	Real-time sequencing	not fixed	not fixed	not fixed	Very long read lengths	Instruments not available at the time of writing
Bench-top instruments						
454 GS Junior + (Roche)	Pyrosequencing	700 bp	18 hours	70 Mb	Long read lengths	Appreciable hands-on time, high reagent costs, high error rate in homopolymers
MiSeq (Illumina)	Reversible terminator	2x300 bp	5-55 hours	0.3-15 Gb	Cost-effectiveness, short run times, appropriate for microbial applications, minimal hands-on time	Read lengths too short for efficient assembly
Ion Personal Genome Machine (Life Technologies)	Proton detection	200 bp	Ion 314: 2.3-3.7 hours/Ion 318: 4.4 - 7.3 hours/Ion 316: 3.0 - 4.9 hours	2 Gb	Short run times, appropriate throughput for microbial applications	Appreciable hands-on time, high error rate in homopolymers
Ion Proton (Life Technologies)	Proton detection	200 bp	2-4 hours	10 Gb	Short run times, flexible chip reagents	Appreciable hands-on time, high error rate in homopolymers

Adapted from (Loman *et al.*, 2012)

Genomics refers to the study of the genomes of organisms. Species with defined genomes increase the value of many different biological studies by enabling the comparison of microbial evolution, metabolism, and ecology of organisms in the environment (DeLong & Karl, 2005). The first sequenced bacterial genome was that of *Haemophilus influenzae* in 1995 (Fleischmann *et al.*, 1995). Since then, the genomes online database (GOLD) (Bernal *et al.*, 2001) lists thousands of bacterial genomes, including both finished and permanent draft genomes. These bacterial genomes vary in terms of G+C content, gene content and density, genome size and number of plasmids. The accumulation of genome sequences is crucial for comparative genome analysis, for at least two related but distinct fundamental reasons: (i) identifying sets of orthologs [*i.e.*, genes that arise by speciation and tend to retain similar functions (Richardson & Watson, 2012)], and (ii) searching for genes that are absent in one but present in another genome sequence. The ability to delineate sets of orthologs and to search for missing genes is indispensable for genome-based reconstruction of an organism's metabolism and other functional systems and for reconstruction of genome evolution.

Bacterial genome sequence analysis involves assembly, gene prediction, and annotation. Genome assembly involves the process of reconstructing a genome sequence by joining short reads together up to the chromosomal level. Assembly procedures can be classified as either reference-guided genome assemblies or *de novo* genome assemblies. For the sake of computational memory saving and convenience of data inquiry, HT short reads data are always initially formatted to a specific data structure. Currently, the majority of existing data structures are graph-based. A well-known graph-based assembly tool is the Velvet *de novo* assembler, which builds a *de Bruijn* graph from short reads and removes errors from the graph (Zerbino & Birney, 2008). Following, gene prediction is the identification of open reading frames encrypted in the DNA sequence. Examples of gene finding tools are Glimmer (Delcher *et al.*, 2007) and GeneMark (Lukashin & Borodovsky, 1998), which are implemented in the rapid annotation using subsystems technology (RAST) (Aziz *et al.*, 2008) and the prokaryotic genome automatic annotation pipeline (PGAAP) (Angiuoli *et al.*, 2008), respectively. The next step is to take the set of predictions and search for hits against one or more protein and/or protein domain databases (*i.e.*, annotation) using BLAST (*i.e.*, basic local alignment tool) (Altschul *et al.*, 1997), HMMer (*i.e.*, hidden markov model) (Eddy, 1998) or other programs. For each gene with a significant match, the BLAST output together with the annotation of the hit can be used to assign a name and function to the corresponding protein. The accuracy of this step depends not only on the annotation software but also on the quality of the annotations already in the reference database. Many genomes of bacteria involved in food fermentations, including several LAB, have already been sequenced [as example (Illeghems *et al.*, 2013; Moreno Switt *et al.*, 2014; Makarova *et al.*, 2006)].

Contrasting to genomics, metagenomics refers to culture-independent studies of the collective set of genomes of mixed microbial populations and applies to explorations of all microbial genomes that reside in a certain niche of interest (Petrosino *et al.*, 2009). It starts with the extraction of genomic DNA from cellular organisms and/or viruses in a sample. This metagenomic DNA can either be sheared into fragments and sequenced with the platform of choice (shotgun sequencing strategy), or one can enrich and sequence a target region of interest (target enrichment strategy) (Figure 1). Metagenomics can be applied to study both qualitative and quantitative aspects of the microbial community of interest. For the qualitative approach, the recognition of the presence of a taxonomic unit or a metabolic function is used to describe the microbial diversity present in the ecosystem. The quantitative approach provides the relative proportion of the taxonomic unit or metabolic function relative to others in the same sample or to other samples or ecosystems. Nevertheless, caution must be taken when interpreting these data because sequencing depth, DNA extraction procedures, PCR biases, and others bias the outcome.

In the field of metagenomics, the shotgun sequencing strategy generates a huge amount of short reads, comprising the metagenome of the community studied. Similarly to genomics, data processing usually involves the assembly of short overlapping sequence reads into a consensus sequence, followed by the prediction and annotation of coding sequences (CDSs). These data enable researchers to look for functionalities of the communities and to identify microbial community members in a process called binning or classification. The metagenomic RAST (MG-RAST) server is an automated analysis platform providing phylogenetic and functional analyses of metagenomes (Meyer *et al.*, 2008). Examples of alternative tools are metaGenome analyzer (MEGAN4) (Huson & Mitra, 2012), integrated microbial genomes/metagenomes (IMG/M) (Markowitz *et al.*, 2014), community cyberinfrastructure for advanced marine microbial ecology research and analysis (CAMERA) (Seshadri *et al.*, 2007), metagenomics Reports (METAREP) (Goll *et al.*, 2010), and others. Food fermentations that have been analyzed using the shotgun sequencing strategy are a kimchi (Jung *et al.*, 2011) and cocoa bean fermentation processes (Illeghems *et al.*, 2012). Despite the increased throughput and resulting reduced prices of HT sequencing technologies, metagenome shotgun sequencing is still expensive and time-consuming, especially for complex metagenomes containing a large number of eukaryotic genomes and for the analysis of a large number of samples. Furthermore, the storage of these large datasets will place a substantial burden on a research center's bioinformatics infrastructure.

Consequently, considerable efforts have been devoted to develop target enrichment methods, in which genomic regions are selectively enriched from a community DNA sample before sequencing (Mamanova *et al.*, 2010). This approach is, compared to the metagenome shotgun sequencing strategy, more cost-effective and the resulting data are considerably less cumbersome to store and to analyze. Its major application is to identify microorganisms in complex communities by exploiting targets

containing both universal and conserved sequences, such as regions within 16S rRNA genes, which enable approximate identification of bacteria and archaea. The same principle holds for the identification of fungal communities by amplifying a fragment of the ITS region. Several tools for target enrichment have been developed, each with their own advantages and disadvantages (Mamanova *et al.*, 2010). For instance, PCR is compatible with any HT sequencing platform, though to make full use of the HT, a large number of amplicons must be sequenced together. However, PCR is difficult to perform in multiplex to any useful degree: the simultaneous use of many primer pairs can generate a high level of nonspecific amplifications, caused by interactions between the primer pairs, and furthermore, amplifications can fail (Cho *et al.*, 1999; Wang *et al.*, 1998). Clever derivatives of multiplex PCR have been developed (Meuzelaar *et al.*, 2007; Varley & Mitra, 2008; Simon *et al.*, 2007), but in practice, it is often more straightforward to perform PCRs in uniplex (Figure 3). Following uniplex amplification, the concentration of PCR products must be normalized before pooling. A convenient solution to many of the problems encountered in a standard PCR-based approach is the RainStorm platform, developed by RainDance Technologies (`http://raindancetech.com/`), which uses microdroplets, similar to emulsion PCR. Each droplet supports an independent PCR and contains a single primer pair along with genomic DNA and other reagents. The entire population of droplets represents many distinct primer pairs and is subjected to thermal cycling, after which this emulsion is broken and PCR products are recovered. The advantages of this technology are that different primer pairs cannot interact with each other, and that direct competition of multiplex PCRs for the same reagent pool is prevented, which should improve uniformity relative to conventional multiplex PCR. Even with an efficient, automatic PCR pipeline, it is not feasible to use conventional PCR to target genomic regions that are several megabases in size because of the high cost of primers and reagents and the DNA input requirements, particularly in large sample sets. Consequently, for very large or moderately sized target regions, other approaches to target enrichment should be used. Alternative strategies involve the usage of molecular inversion probes (Hardenbol *et al.*, 2003) and hybrid capture (Albert *et al.*, 2007).

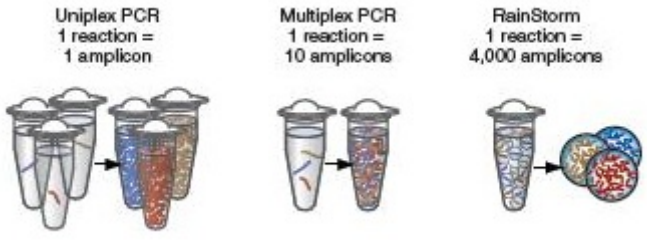


Figure 3: PCR as a target enrichment strategy

(Mamanova *et al.*, 2010)

The use of target enrichment strategies is common in food microbiology. Pyrosequencing of PCR-amplified 16S rRNA genes has been applied to study the bacterial diversity of a number of fermented foods, including narezushi (fermented fish and rice) (Kiyohara *et al.*, 2012; Koyanagi *et al.*, 2011), nukadoko (fermented rice bran) (Sakamoto *et al.*, 2011), Chinese liquor fermentations (Li *et al.*, 2011), Danish raw-milk cheeses (Alegria *et al.*, 2012), Polish oscypek cheese (Alegria *et al.*, 2012), and pearl millet fermentations (Humblot & Guyot, 2009). Additionally, it has been used to analyze the bacterial and archaeal or fungal communities of fermented seafood (Roh *et al.*, 2010), makgeolli (rice beer) (Jung *et al.*, 2012), meju (Kim *et al.*, 2011), doenjang (soybean pastes) (Nam *et al.*, 2012a), kochujang (fermented red pepper condiment) (Nam *et al.*, 2012b), and kimchi (Park *et al.*, 2012). Because of the increasing read length, higher throughput and lower cost of Illumina sequencing, its application recently gained popularity in food microbiology. For instance, Bokulich and coworkers used Illumina sequencing to study botyrized wine fermentations (2012b) and the brewhouses resident microbiota responsible for the fermentation of American coolship ales (ACA) (2012a), beers that are characterized by a spontaneous fermentation process. Compared to Sanger sequencing of PCR clone libraries, these target enrichment strategies permit a much deeper sampling of microbial communities by providing orders of magnitude more sequence information. Moreover, performing PCR in multiplex enables different samples to be analyzed in parallel and makes the inclusion of technical, biological, and experimental replications feasible. These advantages will provide greater sensitivity than PCR clone libraries (Sogin *et al.*, 2006; Parameswaran *et al.*, 2007).

2

Importance of LAB in Food Fermentations

LAB play a prominent role in the world food supply, performing the main bioconversions in fermented dairy products, meats, cereals, and vegetables. They contribute to the taste and texture of fermented food products and inhibit food spoilage bacteria by producing growth-inhibiting substances and large amounts of lactic acid, thereby prolonging the shelf-life of fermented foods.

One of the key metabolic features of LAB is carbohydrate fermentation coupled to substrate-level phosphorylation, finally resulting in the reduction of pyruvate into lactate to regenerate NAD^+ in a process catalyzed by lactate dehydrogenase (Figure 4). Based on differences in carbohydrate fermentation patterns, they are subdivided into homofermentative and heterofermentative strains (Stiles & Holzapfel, 1997). The homofermenters produce lactic acid as the major end-product of carbohydrate fermentation, whereas heterofermenters generate ethanol and CO_2 next to lactic acid. Differences between both categories occur on the enzyme level, with aldolase being absent and phosphoketolase being present in heterofermenters, and vice versa in homofermenters. Aldolase is a key enzyme of the glycolysis, converting fructose 1,6-biphosphate into glyceraldehyde 3-phosphate and dihydroxy acetone phosphate, whereas phosphoketolase converts xylulose 5-phosphate into acetyl phosphate and glyceraldehyde 3-phosphate. The homofermentative metabolism is energetically favored to the heterofermentative one, although the latter has the competitive advantage of catabolizing pentoses next to hexoses,

whereas homofermenters can only utilize hexoses. Next to acidification due to the carbohydrate metabolism, LAB contribute to flavor due to their amino acid catabolism (Figure 4). In general, aminotransferases are involved in the conversion of amino acids (branched-chain amino acids, aromatic amino acids, and sulfur-containing amino acids) into the corresponding α-keto acids, followed by decarboxylation into the corresponding aldehydes and further dehydrogenation into the corresponding alcohols.

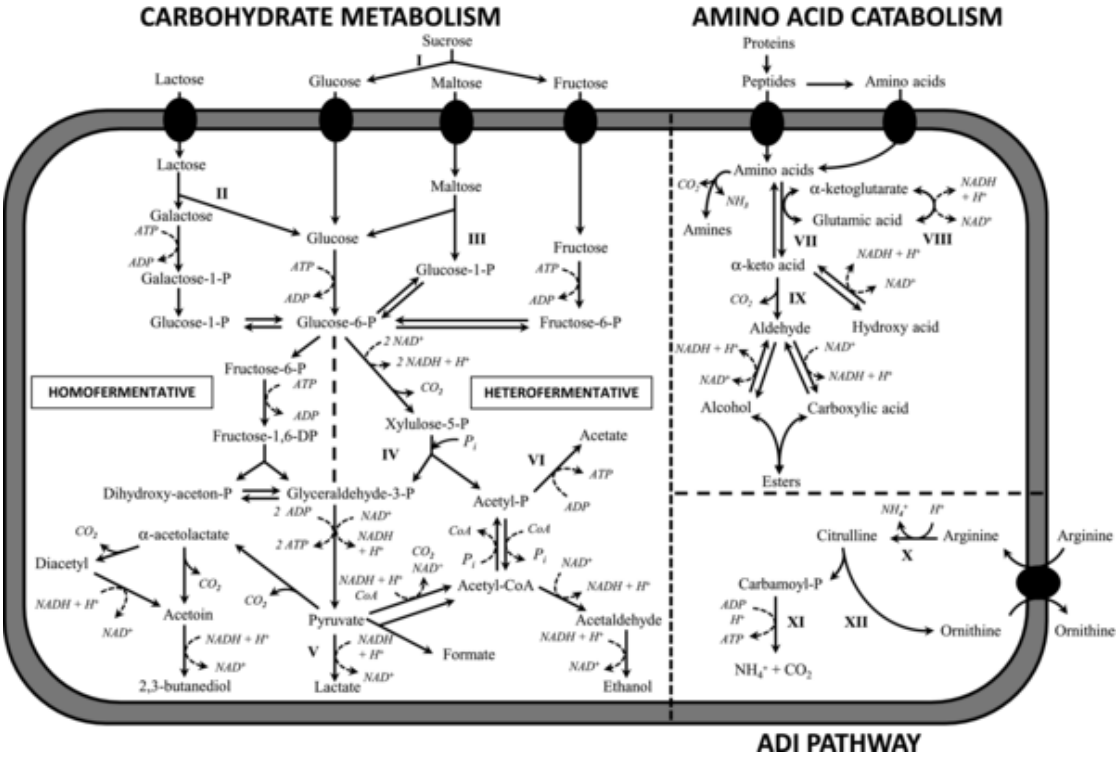


Figure 4: Major pathways active in LAB

Left panel: carbohydrate metabolism via the glycolysis (homofermentative LAB; lower left) and the phosphoketolase pathway (heterofermentative LAB; lower middle). Upper right panel: generic conversion of branched-chain amino acids (valine, leucine, and isoleucine), aromatic amino acids (tyrosine, tryptophan, and phenylalanine), and sulphur-containing amino acids (methionine), as initiated by transamination in LAB. Lower right panel: arginine deiminase pathway (ADI). Key enzymes involved in these pathways are designated with Roman numbers: I, cell wall-bound fructosyltransferase; II, β-galactosidase; III, maltose phosphorylase; IV, phosphoketolase; V, lactate dehydrogenase; VI, acetate kinase; VII, aminotransferase; VIII, glutamate dehydrogenase; IX, decarboxylase; X, arginine deiminase; XI, ornithine transcarbamoylase; XII, carbamate kinase. (Ravyts *et al.*, 2012)

Current taxonomy classifies LAB genera in the order *Lactobacillales* within six families: *Aerococcaceae*, *Carnobacteriaceae*, *Enterococcaceae*, *Lactobacillaceae*, *Leuconostocaceae*, and *Streptococcaceae* (Figure 5). Food-grade LAB are generally restricted to the genera *Lactobacillus*, *Enterococcus*, *Leuconostoc*, *Fructobacillus*, *Weissella*, *Pediococcus*, *Lactococcus*, *Streptococcus*, *Carnobacterium*, *Oenococcus*,

and *Tetragenococcus*. They possess small genomes (approximately 2 Mb), encoding a range of biosynthetic capabilities that reflect both prototrophic and auxotrophic characters (Makarova *et al.*, 2006). Furthermore, their genomes encode a broad repertoire of transporters for efficient carbon and nitrogen acquisition from the nutritionally rich environments they inhabit (Makarova *et al.*, 2006). This suggests both extensive gene loss as well as acquisitions via horizontal gene transfer during the evolution of LAB within their habitats (Makarova *et al.*, 2006). Several strains have plasmids containing genes regulating the fermentation of carbohydrates and encoding resistances. Most LAB are free-living or live in beneficial or harmless associations with animals, although some are opportunistic pathogens.

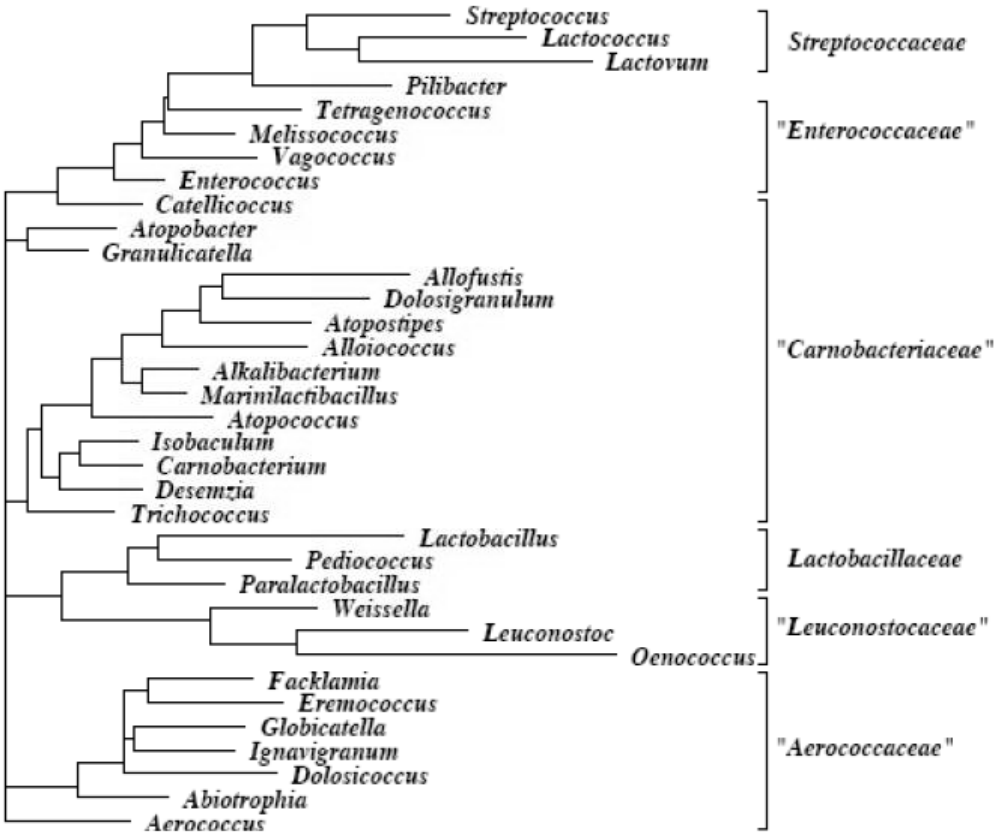


Figure 5: Consensus dendrogram reflecting the phylogenetic relationships of a selection of LAB genera of the order Lactobacillales *within the class* Bacilli

(Garrity *et al.*, 2004). The genus *Leuconostoc* includes *Leuconostoc sensu stricto* and the genus *Fructobacillus* (Endo & Okada, 2008).

Examples of fermented foods, for which LAB are critical in their production are yogurt, cheese, butter, sour cream, sausage, pickles, olives, sauerkraut, cocoa, and others. For instance, LAB play a major role in cocoa bean fermentation, which is characterized by the successional growth of various species of yeasts, LAB, and AAB whereby LAB contribute to the fermentation of carbohydrates and citrate into lactate, acetate, and mannitol (De Vuyst *et al.*, 2010). A diversity of species within the genera *Lactobacillus*, *Streptococcus*, *Lactococcus*, *Oenococcus*, *Pediococcus*, *Weissella*, *Fructobacillus*, and *Leuconostoc* have been isolated from fermented cocoa beans from different locations (Papalexandratou *et al.*, 2011a). LAB also occur in the beer environment, with *Lactobacillus* and *Pediococcus* species being frequently reported as beer spoilers (Bokulich & Bamforth, 2013; Bergsveinson *et al.*, 2012; Pittet *et al.*, 2012). LAB are also known to contribute to the flavor development of certain spontaneously fermented beers, such as lambic beers (Spitaels *et al.*, 2014b) and beers produced by mixed-culture fermentation, such as Belgian red-brown acidic ales, including red acidic ales and red-brown acidic ales (Martens *et al.*, 1997). *Pediococcus damnosus* is a predominant microbial community member during the maturation of these beers, resulting in increased lactic acid concentrations and thereby contributing to the fresh and acidic taste of the bottled beers. Additionally, psychrotrophic LAB are increasingly reported as spoilers within the storage period of cold-stored packaged foods, causing alterations of the organoleptic properties of these food products (Pothakos *et al.*, 2014). Examples of LAB species that successfully proliferate under refrigeration temperatures are *Lactobacillus sakei*, *Lactobacillus curvatus*, *Leuconostoc mesenteroides*, *Leuconostoc gelidum* subsp. *gasicomitatum*, *Leuconostoc gelidum* subsp. *gelidum*, and *Lactococcus piscium* (Pothakos *et al.*, 2014).

An interesting application of LAB is their use as starter cultures, which are defined as microbial preparations of large numbers of cells of at least one microorganism that is added to a raw material to produce a fermented food by accelerating and steering its fermentation process (Leroy & De Vuyst, 2004). The advantages of their use are that a high degree of control over the fermentation process is achieved and standardization of the end-products is reachable. Additionally, LAB can be used as functional starter cultures, which are starters that possess at least one inherent functional property besides lactic acid production. Functional starter cultures can contribute to food safety and/or offer organoleptic, technological, nutritional, or health advantages (Leroy & De Vuyst, 2004; Leroy *et al.*, 2006). Examples are LAB strains that are able to produce antimicrobial substances, carbohydrates, polymers, sweeteners, aromatic compounds, useful enzymes, or nutraceuticals, or LAB with health-promoting properties, so called probiotic strains. Furthermore, progress has been made in the construction of food-grade genetically modified LAB for the improvement of fermented foods (Smid *et al.*, 2006), and for usage as mucosal delivery vectors for therapeutic proteins and DNA vaccines (Bermúdez-Humarán *et al.*, 2013; Sybesma *et al.*, 2006).

3

LAB Taxonomy

The availability of an appropriate taxonomic framework in sequence databases is crucial for organizing and cataloging microbial diversity (Yilmaz *et al.*, 2013). Taxonomy relies on three key elements depending on each other: classification (*i.e.*, the orderly arrangement of organisms into taxonomic groups on the basis of similarity), nomenclature (*i.e.*, the labeling of taxonomic groups), and identification of unknowns (*i.e.*, the process of determining whether an organism belongs to one of the taxonomic groups or represents a novel taxon). While nomenclature is governed by the International Code of Nomenclature of Bacteria (1990 Revision) (Lapage *et al.*, 1992), the classification and identification of prokaryotes has been a changing area in the last 50 years. Phylogeny elucidates evolutionary relationships among organisms and is a tool for classification. The 16S rRNA gene is one of the best targets for phylogenetic studies of bacteria and its sequences therefore provide a backbone in taxonomy, as they yield a bacterial phylogenetic framework (Vandamme *et al.*, 1996). The bacterial species is an unit in biological classification, which is crucial for the construction of a useful taxonomic framework based on microbial evolution (Gevers *et al.*, 2005; Dijkshoorn *et al.*, 2000). Various species concepts, describing bacterial species and explaining how they are formed, have been suggested for microbial species (see below), but none have been generally accepted (Rossello-Mora & Amann, 2001). The definition of a bacterial species is how the concept is applied in practice and consists of a set of rules (see box below) (Konstantinidis *et al.*, 2006). While the definition of an eukaryotic species has focused on the biology of reproduction, the bacterial species has mainly been discussed in terms of similarity between strains.

Species concepts in bacteriology (Achtman & Wagner, 2008)
'A species could be described as a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity in many independent characteristics, and is diagnosable by a discriminative phenotypic property (Rossello-Mora & Amann, 2001).'
'Species are considered to be an irreducible cluster of organisms diagnosably different from other clusters and within which there is a parental pattern of ancestry and descent (Staley, 2006).'
'A species is a group of individuals where the observed lateral gene transfer within the group is much greater than the transfer between groups (Dykhuizen, 2005).'
'Microbes do not form natural clusters to which the term 'species' can be universally and sensibly applied (Nesbo *et al.*, 2006).'
'Species are (segments of) metapopulation lineages (de Queiroz, 2005).'

3.1 Phylogenetics

The term 'phylogenetics' refers to the study of the evolutionary history of species, organisms, genes, or proteins through the construction and analysis of mathematical entities known as trees or phylogenies (Sleator, 2011). It involves the alignment of nucleic acid or protein sequences between extant organisms. An inference is then generated to explain the repartition of character states and the results are presented as a phylogenetic tree, which is a graphic representation of the computed results (Sleator, 2011; Hall, 2011).

While several methods exist to infer evolutionary relatedness, most can be classified as either distance- or character-based methods (Figure 6). Distance-based methods utilize an algorithm, incorporating a model of evolution, to compute a distance-matrix from which a phylogenetic tree is calculated by means of progressive clustering (Sleator, 2011). Distances in the matrix relate to the number of differences between each pair of DNA or protein sequences. The tree is constructed from the numerical data in the matrix, with the most closely related sequences occupying a position on the tree which is distant from the less closely related sequences. Both the neighbour-joining (NJ) and the unweighted pair group method using arithmetic averages (UPGMA) approaches employ distance-based methods (Figure 6). Although effective, distance-based methods have a number of disadvantages. For instance, NJ may compute different trees depending on the order in which the constituent sequences are added (Sleator, 2011). Furthermore, NJ provides only a single tree rather than a consensus tree (Sleator, 2011). Character-based methods, such as maximum parsimony (MP) and maximum likelihood (ML) methods, search for the most probable tree for a specific sequence set, based on characters at each

position of the sequence alignment (Sleator, 2011). Each character is considered one at a time to calculate the 'tree score', representing the minimum number of changes for maximum parsimony and the log-likelihood value for maximum likelihood. *MP* seeks to find the tree(s) that are compatible with the minimum number of substitutions among sequences, *i.e.*, the fewest evolutionary changes. An advantage of MP is that it provides diagnosable units for each clade and branch lengths in terms of the number of changes on each branch. However, a significant limitation of the MP approach is that it requires strict assumptions of consistency across sites and among lineages. *ML* methods are based on specific probabilistic models of evolution and search for the tree with maximum likelihood under these models. The major advantage of likelihood approaches is that they are based on powerful statistical theory. However, ML approaches are computationally intensive, limiting its use to a relatively small number of sequences. The distance-based and ML methods make use of a model of evolution and as such are model based, while MP does not have an explicit model.

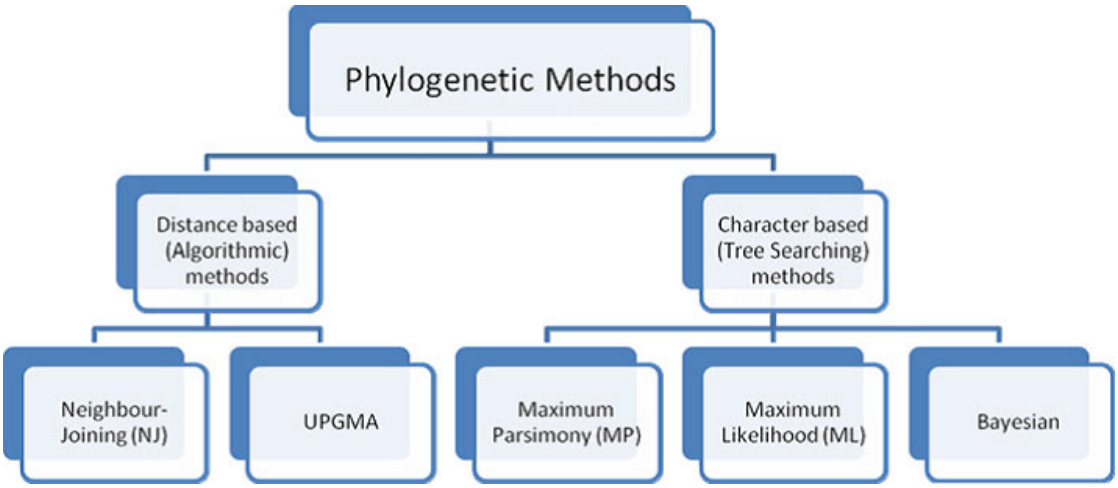


Figure 6: Schematic overview of the major analytical approaches to phylogenetic tree building.

(Sleator, 2011). The Bayesian inference approach is not discussed in the text.

In order te reconstruct evolutionary trees, assumptions about the substitution process must be made by means of a model of evolution. A model of evolution makes assumptions on how DNA or amino acid substitutions occurred in the DNA or protein sequence since they last shared a common ancestor (Sleator, 2011; Hall, 2011). Because multiple substitutions can occur at the same site, there is a potential for underestimating the amount of change that has occurred along a branch. Each model of evolution represents an attempt to account for multiple hits and various other aspects of the substitution process. The easiest model to consider is one in which the probabilities of any nucleotide changing to any other nucleotide

are equal (*i.e.*, One-Parameter or Jukes-Cantor model. The One-Parameter Model is the simplest, but it is not very realistic because all changes do not occur at the same rates. Therefore, a variety of models have been proposed that allow to specify different rates. The most general model is one in which each different substitution can occur at a different rate, involving 12 free-parameters. A lot of important models are special cases of this general model. For instance, the Kimura 2-parameter model (Kimura, 1980) extends the Jukes-Cantor correction by taking into account the possibility that the rates at which transitions and transversions occur might be different. The Tamura 3-parameter model adds a correction for compositional bias (Tamura, 1992). The general time-reversible (GTR) model has six different substitution probabilities. It involves nine free parameters and the equilibrium base frequencies. The substitution process looks the same whether we observe the process with time running forward or backward (Hall, 2011).

The models listed above implicitly assume that the rates are the same at all sites, but it is also possible to include rate variation across sites in the models. An example of site-specific variation is that second positions in codons evolve most slowly, first positions evolve at an intermediate rate, and third positions evolve most rapidly. Furthermore, because proteins fold, those slowly evolving sites tend to be located in patches rather than in one specific region of the sequence. Estimating these individual rates for each site is often computationally impractical when many sequences are involved. Therefore, the alternative is to assume a well-behaved distribution of rates across sites. The distribution that is commonly used is the gamma distribution. Furthermore, some sites, such as initiation codons, may not be free to vary at all. This constitutes a specific case of site-specific variation, which we call the invariant sites (Hall, 2011).

Once an optimum tree is chosen, some statistical measure of internal support for clades must be obtained to ascertain whether the tree is sufficiently robust and biologically meaningful (Sleator, 2011). A method that is frequently used to verify the evolutionary reliability of trees is the bootstrap analysis that pseudoreplicates the collection of data as a method to estimate the reliability of the tree. A random site is taken from the alignment and is used as the first site in a pseudoalignment. Another random site is taken and used as the second site in the pseudoalignment, and this process is continued until the pseudoalignment contains the same number of sites as the original alignment. The tree is then constructed from the pseudoalignment using the same method and under the same parameter settings used to estimate the original tree (Hall, 2011).

3.2 Polyphasic Taxonomy

Developments in the field of sequencing of rRNA genes contributed to bacterial phylogeny and molecular fingerprinting techniques and resulted in the introduction of the polyphasic taxonomy concept (Vandamme *et al.*, 1996). Polyphasic taxonomy comprises the determination of the phylogenetic position of an isolate, based on measures of evolutionary relationships, which are in turn based on gene sequences (*e.g.*, the 16S rRNA gene) and phenotypic properties to assess novelty. Within this framework of polyphasic taxonomy, strains belonging to the same species should form differentiable coherent groups, displaying similar phenotypes, genotypes, and chemotaxonomic features. The polyphasic taxonomy scheme defines a bacterial species as a group of strains having a certain degree of phenotypic consistency, $>$ 70% DNA-DNA hybridization (DDH) similarity, and $>$ 97% 16S rRNA gene sequence identity (Vandamme *et al.*, 1996).

In the case of LAB species, unraveling phylogenetic relationships is cumbersome because of the low taxonomic resolution of the 16S rRNA gene, resulting in only an approximate or tentative identification if high ($>$ 97%) similarity values are found. Therefore, MLSA of different housekeeping genes is often needed to obtain a species level identification. The first comprehensive effort to make a large database for the classification and identification of LAB has been made by Naser and coworkers (2005a; 2005b; 2007). They evaluated the MLSA of three housekeeping genes [*i.e.*, *pheS* encoding phenylalanine transfer RNA (tRNA) synthetase, *rpoA* encoding RNA polymerase α-chain, and *atpA* encoding RNA polymerase α-subunit] as species identification tool within the genera *Enterococcus* and *Streptococcus*, and two of these genes (*i.e.*, *pheS* and *rpoA*) within the genus *Lactobacillus* (Naser *et al.*, 2005a). This MLSA scheme was expanded by De Bruyne and coworkers (2007), who constructed a MLSA scheme based on *pheS*, *rpoA*, and *atpA* for the genera *Pediococcus*, *Leuconostoc*, and *Fructobacillus*. Furthermore, a single-locus sequence analysis framework was established based on *pheS* for the genera *Weissella* and *Oenococcus* (De Bruyne *et al.*, 2008a). Furthermore, Rademaker and coworkers (2007) extended this list with MLSA for the genus *Lactococcus* based on the *pheS*, *rpoA*, *atpA*, *bcaT* (encoding branched chain amino acid aminotransferase), *pepN* (encoding aminopeptidase N), and *pepX* (encoding X-prolyl dipeptidyl peptidase) gene sequences.

If a LAB strain or a set of strains belong to a novel taxon, they should be characterized as comprehensive as possible (Tindall *et al.*, 2010). The goal of this characterization is to place them within the hierarchical framework laid down by the International Code of Nomenclature of Bacteria (Lapage *et al.*, 1992), as well as to provide a description of the novel taxon. The starting point for novel LAB species description is often the availability of a full-length 16S rRNA gene sequence with less than 97% sequence similarity towards its closest neighbor

(Stackebrandt & Goebel, 1994). When the 16S rRNA gene similarity exceeds 97%, MLSA and/or DDH are needed to determine whether the strain represents a novel taxon (Mattarelli *et al.*, 2014). For LAB species, the phenotypic description should also include the composition of the peptidoglycan of the cell wall and the ratio of D- to L-lactic acid production, as these are species-specific (Mattarelli *et al.*, 2014).

3.3 Genomic Taxonomy

Despite the widely used phylogenetic analysis based on the 16S rRNA gene, there has been considerable debate whether a tree based on any single gene can accurately represent the evolution of a species. Hence, the importance of the frequency of recombination, either in the form of homologous recombination (Fraser *et al.*, 2007) or lateral gene transfer (Doolittle & Papke, 2006), has challenged bacterial species concepts and raised many questions. Still, it has been shown that these forms of recombination are considered cohesive rather then disruptive forces in bacterial species (Konstantinidis & Tiedje, 2005). Furthermore, DDH experiments have several disadvantages, including the requirement of large quantities of high-quality DNA, its time-consuming and labor-intensive nature, and the inability to build databases (Gevers *et al.*, 2005). Because of these drawbacks, bacterial taxonomists are actively searching for alternative methods that can abandon DDH experiments (Gevers *et al.*, 2005; Coenye *et al.*, 2005; Cho & Tiedje, 2001).

DDH experiments were originally developed because the *ad hoc* committee on reconciliation of approaches to bacterial systematics (Wayne *et al.*, 1987) stated that taxonomy should be determined phylogenetically and that the complete genome sequence should therefore be the standard for species delineation. Particularly interesting within this regard is development of HT sequencing technologies, which caused the accumulation of whole-genome sequence data and allowed the establishment of taxonomic schemes based on evolutionary information contained in whole-genome sequences (Eisen, 2000; Wolf *et al.*, 2001).

Genomic taxonomy integrates comparative genomics in taxonomy and can include MLSA, phylogenetic analysis based on whole-genome sequences, average nucleotide identity (ANI) (Konstantinidis & Tiedje, 2005; Richter & Rossello-Mora, 2009), average amino acid identity (AAI), the percentage of conserved DNA (pcDNA) (Goris *et al.*, 2007), maximal unique matches (MUM) indices (Deloger *et al.*, 2009), core and pan genome analysis, genomic signatures, codon usage bias, metabolic pathway content, and others (Konstantinidis & Stackebrandt, 2013). Phylogenetic analysis based on whole-genome sequences are based on all genes of the core genome. The ANI and AAI are calculated among the nucleotide or amino acid sequences, respectively, of the conserved genes or DNA of a pair of genomes

(Konstantinidis & Tiedje, 2005; Goris *et al.*, 2007). Furthermore, the calculation of the pcDNA or MUM indices can be used as *in silico* imitations of DDH experiments, whereas the analysis of the ANI and the core genome are both natural expansions of MLSA. ANI represents a robust measure of the genetic and evolutionary distances between two sequenced strains, because it shows a strong correlation with 16S rRNA gene sequence similarity and the mutation rate of the genome, it is not affected by lateral gene transfer or variable recombination rates of single (or a few) genes, and it offers resolution at the subspecies level (Konstantinidis & Tiedje, 2005). The recommended cut-off of 70% DDH for species delineation corresponds to 95% ANI (Goris *et al.*, 2007). Finally, genomic signatures are compositional parameters, reflecting the di-, tri-, or tetranucleotide relative abundances (Karlin signature), which are similar between closely related species and dissimilar between non-related species.

The continuing accumulation of LAB genome sequences provides a good opportunity to study the evolutionary history of LAB species. The first LAB genome sequence was published in 2001, *i.e.*, that of *Lactococcus lactis* subsp. *lactis* IL 1403, a strain commonly used as cheese starter (Bolotin *et al.*, 2001). Nowadays, numerous LAB genome sequences are available, mainly belonging to the genera *Lactobacillus* and *Streptococcus*. Phylogenomic studies on LAB provide interesting new information but thus far include a rather limited number of species (Coenye & Vandamme, 2003; Makarova *et al.*, 2006; Zhang *et al.*, 2011; Claesson *et al.*, 2008).

Coenye and Vandamme (2003) analyzed 10 LAB genome sequences belonging to the genera *Streptococcus* (8 genomes), *Lactococcus* (1 genome), and *Lactobacillus* (1 genome). They showed that trees based on different kinds of phylogenetic information extracted from these genomes do not provide much additional information about the phylogenetic relationships among LAB taxa compared to more traditional alignment-based methods. Nevertheless, they expected that the study of the genome sequences would have their value in taxonomy by determining which genes are shared and when genes or sets of genes have been lost during evolution. Furthermore, these genome sequences could aid in detecting the presence of horizontally transferred genes and/or confirming or enhancing the phylogenetic signal derived from traditional methods. Next, Makarova and colleagues (2006) compared 12 complete genome sequences, covering the major families within the order *Lactobacillales* [*Leuconostocaceae* (2 genomes), *Lactobacillaceae* (7 genomes), and *Streptococcaceae* (3 genomes)]. Phylogenetic analysis of multiple protein CDSs showed that the streptococci-lactococci branch is basal in the *Lactobacillales* tree and that the *Pediococcus* group is a sister of the *Leuconostoc* group, which supports the paraphyly of the *Lactobacillus* genus. Lefébure and Stanhope (2007) applied comparative evolutionary analysis on 26 *Streptococcus* genomes to assess levels of recombination and positive selection in pathogen adaptation to their hosts. This study indicates that positive selection appears to be of main importance in species differentiation and adaptation to new hosts, whereas the process of recombination

has a role in strain evolution. Claesson and coworkers (2008) conducted a phylogenomic study on 12 *Lactobacillus* genomes and observed incongruence between the tree based on 141 core proteins and single-gene trees. The source of incongruence is so far unknown, but is likely due to different evolutionary rates among the genes, hidden paralogies, or horizontal gene transfer. Finally, Zhang and coworkers (2011) studied the phylogenetic relationships among 28 LAB (comprising 7 genera of 4 families) genome sequences based on 232 core genes. The concatenation of these genes allowed the recovery of a strongly supported phylogeny, providing a maximum and decisive resolution of the relationships among the LAB species examined.

Part III

Experimental Work

4

Traditional Workflow in Microbial Diversity Analyses

Traditional methods for microbial diversity analyses were applied to two food ecosystems. The first ecosystem studied was the spontaneous cocoa bean fermentation ecosystem. Cocoa beans, which are the seeds of the fruit pods of the cocoa tree *Theobroma cacao*, undergo a fermentation prior to further processing. In the pods, the cocoa beans are embedded in a mass of mucilaginous white pulp, consisting of pectin, citric acid, and carbohydrates. After removal of both beans and pulp from the pods, they are piled into heaps, covered with plantain leaves, and a spontaneous fermentation starts. A succession of microbial activities takes place, encompassing a restricted species diversity of yeasts, LAB, and AAB. An isolation campaign was carried out by Papalexandratou and coworkers (2011a; 2011c; 2011d) from spontaneously cocoa bean fermentations carried out in Brasil, Ecuador, and Malaysia. The resulting isolates were subsequently dereplicated using rep-PCR fingerprinting. This dereplication gained a set of isolates that could not be identified on the species level. Therefore, the polyphasic taxonomy approach was applied to these isolates in Section 4.1, which led to the description of *Weissella fabalis* sp. nov. and the identification of several *Fructobacillus tropaeoli* strains.

The second ecosystem studied was rutabaga (*Brassica napobrassica*), which is a cruciferous plant with a thick bulbous edible yellow root. The edible yellow root was sous-vide cooked, packaged, and stored. Sous-vide cooking or vacuum

cooking is defined as: 'raw materials or raw materials with intermediate foods that are cooked under controlled conditions of temperature and time inside heat-stable vacuumized pouches'. Lower temperatures (usually under 100°C) and longer cooking times than those used in traditional cooking are typically applied. A Finnish laboratory isolated a set of strains from sous-vide cooked rutabaga. These isolates were dereplicated by means of MALDI-TOF MS and resulted in the description of a novel *Leuconostoc rapi* sp. nov. using the polyphasic taxonomy approach (Section 4.2).

Finally, Peeters and coworkers (2011) obtained several Antarctic isolates using plate culturing and subsequent dereplication using rep-PCR. This resulted in a set of unidentified isolates, which were subsequently identified in Section 4.3 as a novel species of the genus *Carnobacterium* (*i.e.*, *Carnobacterium iners* sp. nov.) using the polyphasic taxonomy approach. Furthermore, this novel species description was accompanied by the construction of an MLSA scheme for the genus *Carnobacterium*, which includes food-grade LAB species.

4.1 Characterization of Strains of *Weissella fabalis* sp. nov. and *Fructobacillus tropaeoli* from Spontaneous Cocoa Bean Fermentations

I. Snauwaert, Z. Papalexandratou, L. De Vuyst, P. Vandamme

Redrafted from: International Journal of Systematic and Evolutionary Microbiology, 2013 (63), 1709-16.

Author contributions: IS performed the experiments and wrote the manuscript. ZP did the isolation and dereplication of the strains. PV and IS designed the experiment. LDV and PV proofread the manuscript.

Six facultatively anaerobic, non-motile LAB were isolated from spontaneous cocoa bean fermentations carried out in Brazil, Ecuador, and Malaysia. Phylogenetic analysis revealed that one of these strains, designated M75^T, isolated from a Brazilian cocoa bean fermentation, had the highest 16S rRNA gene sequence similarity towards *Weissella fabaria* LMG 24289^T (97.7%), *Weissella ghanensis* LMG 24286^T (93.3%), and *Weissella beninensis* LMG 25373^T (93.4%). The remaining LAB isolates, represented by strain M622, showed the highest 16S rRNA gene sequence similarity towards the type strain (T) of *Fructobacillus tropaeoli* (99.9%), a recently described species isolated from a flower in South Africa. *pheS* gene sequence analysis indicated that the former strain represented a novel species, whereas *pheS*, *rpoA*, and *atpA* gene sequence analysis indicated that the remaining five strains belonged to *F. tropaeoli*; these results were confirmed by DDH experiments towards their respective nearest phylogenetic neighbors. Additionally, MALDI-TOF MS proved successful for the identification of species of the genera *Weissella* and *Fructobacillus* and for the recognition of the novel species. We propose to classify strain M75^T (=LMG 26217^T =CCUG 61472^T) as the type strain of the novel species *Weissella fabalis* sp. nov.

4.1.1 Introduction

Papalexandratou and colleagues (2011a; 2011c; 2011d) analyzed fermenting cocoa pulp-bean mass samples collected from traditional Brazilian (October-November 2006), Ecuadorian (April 2008), and Malaysian (April 2010) cocoa bean box and platform fermentations. In these studies, they isolated a large number of LAB that were first screened and identified by $(GTG)_5$ -PCR fingerprinting. However, this approach failed to identify one isolate, designated M75^T, from a Brazilian cocoa bean box fermentation, with a unique $(GTG)_5$-PCR fingerprint and another 63 LAB isolates (two, 58 and three isolates from Brazil, Ecuador, and Malaysia, respectively) with similar $(GTG)_5$-PCR fingerprints. The precise taxonomic positions of six of these LAB, which belonged to five different rep clusters [clusters I-III, M75^T, M56, and M622 (Brazil); cluster IV, M1588 and M1190 (Malaysia and Ecuador, respectively); cluster V, M710 (Ecuador)] were determined in the present study.

4.1.2 Methods, Results, and Discussion

Analysis of the 16S rRNA gene sequences of strains M75^T and M622 was performed as described by De Bruyne *et al.* (2008a), except that sequencing reactions were purified using the BigDye xTerminator purification kit according to the protocol of the supplier (Applied Biosystems). The ARB software package (Ludwig *et al.*, 2004) and the corresponding SILVA SSURef 102 database (Pruesse *et al.*, 2007) were used to align the 16S rRNA gene sequences obtained and those of the type strains of all established species of the genera *Fructobacillus* and *Weissella*, their nearest phylogenetic neighbors (see below). These aligned sequences were imported into the software package molecular evolutionary genetics analysis (MEGA) version 5.0 (Tamura *et al.*, 2011) and analyzed using the neighbor-joining, maximum-likelihood (ML), and maximum-parsimony methods. The statistical reliability of tree topologies was evaluated by bootstrapping analysis based on 1000 tree replicates. The maximum-parsimony and neighbor-joining trees (not shown) revealed topologies similar to those obtained in the phylogenetic tree reconstructed using the ML approach (Figure 7). Sequence similarity calculations performed using the ARB software package indicated that the closest relatives of strain M75^T were *W. fabaria* LMG 24289^T (97.7%), *W. ghanensis* LMG 24286^T (93.3%), and *W. beninensis* 2L24P13^T (93.4%). Lower sequence similarities ($<$ 89%) were found towards other species of the genus *Weissella* with validly published names. Strain M622 was most closely related to *F. tropaeoli* F214-1^T (99.9%) and *F. pseudoficulneus* LC 51^T (99.2%). Lower sequence similarities ($<$ 98%) were found towards other species of the genus *Fructobacillus* with validly published names.

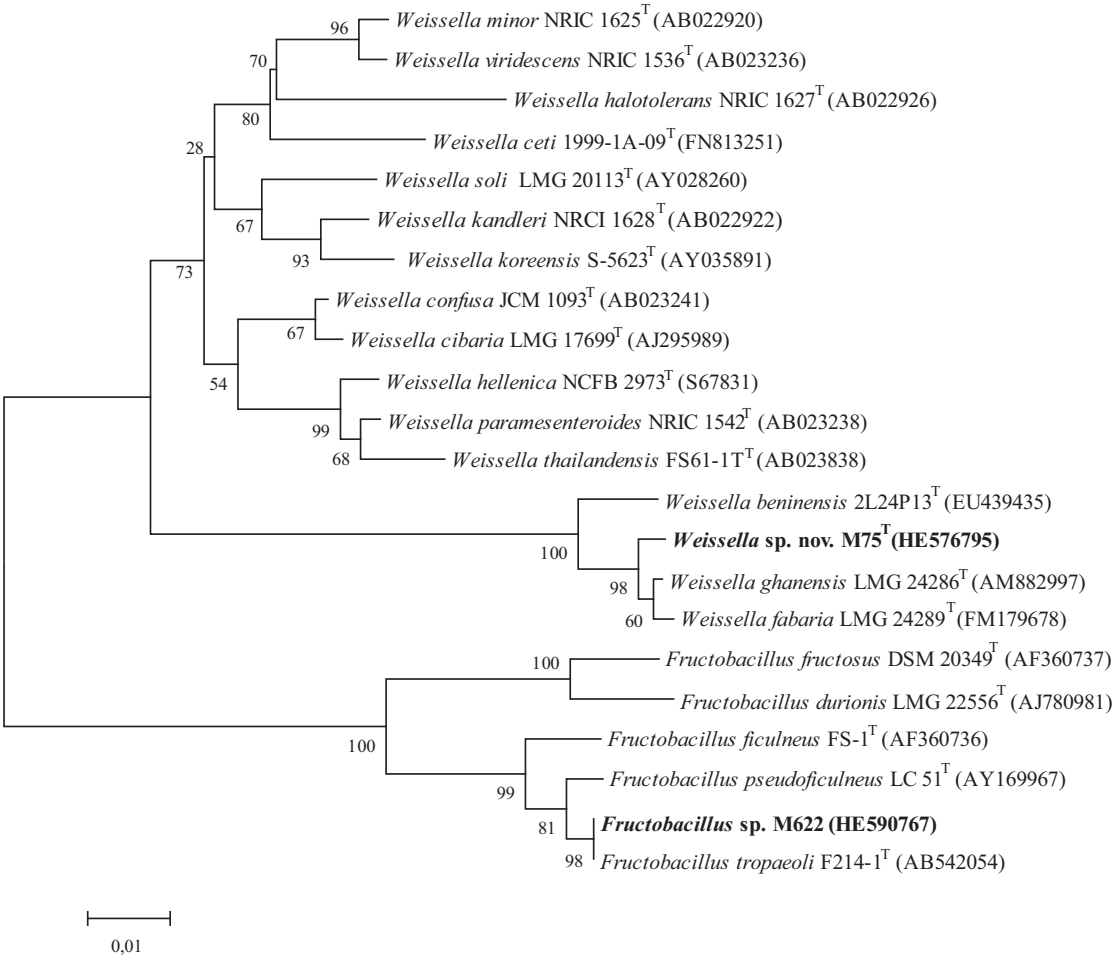


Figure 7: ML tree based on 16S rRNA gene sequences showing the phylogenetic relationships of strains M75^T and M622 among the type strains of all species of the genera Weissella *and* Fructobacillus

Bootstrap values (%) based on 1000 replications are shown at branch points. The substitution model used was the general time reversible (GTR) model and the aligned sequence had a length of 1345 bp. Bar, 1% sequence divergence.

Sequence analysis of the housekeeping genes *pheS*, *rpoA*, and *atpA* correlates with species delineation as determined by DDH in the genus *Fructobacillus* (De Bruyne *et al.*, 2007). Therefore, *pheS*, *rpoA*, and *atpA* gene sequences of strains M56, M622, M710, M1190, and M1588 and *F. tropaeoli* LMG 26298^T, for which no *pheS*, *rpoA*, or *atpA* sequences were available, were determined as described previously (De Bruyne *et al.*, 2007), except that sequencing reactions were purified using

the BigDye xTerminator purification kit as described above. Similarly, sequence analysis of the *pheS* gene has been used successfully to distinguish all established species of the genus *Weissella* (De Bruyne *et al.*, 2008a,0). Therefore, *pheS* gene sequences were determined for strain M75^T and for *W. beninensis* LMG 25373^T and *W. fabaria* M1160 and M1167, for which no *pheS* gene sequences were available. *pheS*, *rpoA*, and *atpA* gene sequences of all remaining *Fructobacillus* and *Weissella* reference strains were available from previous studies (De Bruyne *et al.*, 2007). Details of strains, depositors, and accession numbers are given in Table 2. SeaView version 4 was used to concatenate the *pheS*, *rpoA*, and *atpA* gene sequences (Gouy *et al.*, 2010) and the software package MEGA version 5.0 (Tamura *et al.*, 2011) was used to align the translated concatenated gene sequences and to analyze the nucleotide sequences as mentioned above.

The maximum-parsimony and neighbor-joining trees (not shown) revealed topologies similar to those obtained in the phylogenetic tree reconstructed using the ML approach (Figures 8 and 9) for both analyses. The concatenated *pheS*, *rpoA*, and *atpA* gene sequences of strains M56, M622, M710, M1190, and M1588 revealed high similarity to the concatenated sequence of *F. tropaeoli* LMG 26298^T (98.4, 97.4, 99.2, 99.3, and 99.3% similarity, respectively). Lower gene sequence similarities were found ($<$ 95%) towards other species of the genus *Fructobacillus* with validly published names. The phylogenetic tree of the concatenated *pheS*, *rpoA*, and *atpA* gene sequences confirmed the discriminatory power of these sequences for species identification within the genus *Fructobacillus*: all species were clearly delineated above 97% concatenated gene sequence similarity (Figure 8). Based on these results and the previously established correlation between MLSA of concatenated *pheS*, *rpoA*, and *atpA* gene sequences and levels of DDH, we concluded that strains M56, M622, M710, M1190, and M1588 belong to *F. tropaeoli*. In addition, pairwise *pheS* gene sequence similarity calculations, calculated using the MEGA software version 5.0, confirmed that the closest relatives of strain M75^T were *W. fabaria* LMG 24289^T (85.2%), *W. ghanensis* LMG 24286^T (86.9%), and *W. beninensis* LMG 25373^T (80%). Lower *pheS* gene sequence similarities ($<$ 78%) were found towards other species of the genus *Weissella* with validly published names. With the exception of *Weissella viridescens*, De Bruyne *et al.* (2010; 2008a) demonstrated that strains belonging to the same species of the genus *Weissella* share *pheS* gene sequence similarity of at least 96.8%. Therefore, the present results suggest that strain M75^T represents a novel species of the genus *Weissella* (Figure 9).

Table 2: Weissella *and* Fructobacillus *strains included in the phylogenetic study and in the MALDI-TOF MS analysis.*

Strain	Depositor	Source	Accession numbers		
			pheS	*rpoA*	*atpA*
W. beninensis					
LMG 25373^T	SW. Padonou	Submerged fermented cassava (Benin)	HE576797		
W. cibaria					
LMG 13587	L. Devriese	Dog, ear (Belgium)	FM202101		
LMG 17699^T	G. Rusul	Chili bo (Malaysia)	FM202102		
LMG 17704	G. Rusul	Chili bo (Malaysia)	FM202100		
LMG 18507	NCFB	Post-harvest deterioration of sugar cane	FM202098		
LMG 21843	KCTC	Partially fermented kimchi (Korea)	FM202099		
R-31690	CCMM	Fermented skimmed milk (Morocco)	FM202103		
W. confusa					
LMG 9497^T	A. Ledeboer	Sugar cane	FM202105		
LMG 11983	ATCC	Grass silage	FM202104		
LMG 14040	L. Devriese	Dog, ear (Belgium)	FM202107		
LMG 17718	G. Rusul	Chili bo (Malaysia)	FM202108		
LMG 18480	G. Rusul	Tapai (Malaysia)	FM202106		
W. fabaria					
LMG 24289^T	Own isolate	Fermenting cocoa pulp-bean mass (Ghana)	FM202097		
252 (R-34084)	Own isolate	Fermenting cocoa pulp-bean mass (Ghana)	FM202096		
M1160 (R-46910)	Own isolate	Fermenting cocoa pulp-bean mass (Ecuador)	HE577177		
M1167 (R-46912)	Own isolate	Fermenting cocoa pulp-bean mass (Ecuador)	HE577176		
W. ghanensis					
LMG 24286^T	Own isolate	Fermenting cocoa pulp-bean mass (Ghana)	FM202095		
194B (R-27442)	Own isolate	Fermenting cocoa pulp-bean mass (Ghana)	FM202094		
W. halotolerans					
LMG 9469^T	DSMZ	Sausage	FM202114		
W. hellenica					
LMG 15125^T	NCFB	Naturally fermented sausage (Greece)	FM202110		
W. kandleri					
LMG 18979^T	NCIMB	Desert spring (Namibia)	FM202116		
W. koreensis					
LMG 21853^T	KCTC	Kimchi (Korea)	FM202115		
W. minor					
LMG 9847^T	NCFB	Slime from milking machine	FM202117		
W. paramesenteroides					
LMG 9852^T	NCFB	Fermented dry salami	FM202111		
W. soli					
LMG 20113^T	S. Roos	Garden soil (Sweden)	FM202113		
LMG 20114	S. Roos	Garden soil (Sweden)	FM202112		
W. thailandensis					
LMG 19821^T	JCM	Fermented fish (Thailand)	FM202109		

Continued on next page

Table 2 – continued from previous page

Strain	Depositor	Source	Accession numbers		
			pheS	*rpoA*	*atpA*
W. viridescens					
LMG 3507^T	NCFB	Cured meat products	FM202120		
LMG 11497	NCFB	NK	FM202122		
LMG 12021	JCM	NK	FM202121		
LMG 13093	NCFB	Frankfurters	FM202119		
LMG 23120	J. Björkroth	Spanish blood sausage (Spain)	FM202118		
***Weissella.* sp. nov.**					
M75^T (LMG 26217^T)	Own isolate	Fermenting cocoa pulp-bean mass (Brazil)	HE576796		
F. durionis					
LMG 22556^T	J. Leisner	Tempoyak made from durian fruit (Malaysia)	AM711166	AM711309	AM711205
LMG 22557	J. Leisner	Tempoyak made from durian fruit (Malaysia)	AM711140	AM711290	AM711171
LMG 22558	J. Leisner	Tempoyak made from durian fruit (Malaysia)	AM711141	AM711291	AM711172
F. ficulneus					
LMG 21928^T	DSMZ	Ripe fig (Portugal)	AM711151	AM711342	AM711183
F. fructosus					
LMG 9498^T	A.Ledeboer	Flower (Japan)	AM711194	AM711321	AM711174
F. pseudoficulneus					
LMG 23899^T	CECT	Ripe fig (Portugal)	AM711281	AM711355	AM711274
R-35156	R. Tenreiro	Ripe fig (Portugal)	AM711235	AM711335	AM711259
R-35157	R. Tenreiro	Ripe fig (Portugal)	AM711236	AM711336	AM711260
R-35158	R. Tenreiro	Ripe fig (Portugal)	AM711237	AM711337	AM711261
R-35159	R. Tenreiro	Ripe fig (Portugal)	AM711238	AM711338	AM711262
R-35160	R. Tenreiro	Ripe fig (Portugal)	AM711239	AM711333	AM711263
F. tropaeoli					
LMG 26298^T	L. Dicks	Flower (South Africa)	HE590678	HE590679	HE590680
M56 (R-46388)	Own isolate	Fermenting cocoa pulp-bean mass (Brazil)	HE590681	HE590682	HE590683
M622 (R-46389)	Own isolate	Fermenting cocoa pulp-bean mass (Brazil)	HE590684	HE590685	HE590686
M710 (R-46397)	Own isolate	Fermenting cocoa pulp-bean mass (Brazil)	HE590687	HE590688	HE590689
M1190 (R-46399)	Own isolate	Fermenting cocoa pulp-bean mass (Brazil)	HE590690	HE590691	HE590692
M1588 (R-46401)	Own isolate	Fermenting cocoa pulp-bean mass (Brazil)	HE590693	HE590694	HE590695

Depositors listed in Table 2: S. W. Padonou, Université d'Abomey-Calavi, Cotonou, Benin; L. Devriese, Ghent University, Ghent, Belgium; G. Rusul, University Putra, Malaysia; NCFB, National Collection of Food Bacteria (now NCIMB), Aberdeen, UK; KCTC, Korean Collection for Type Cultures, Yusong, Taejon, Republic of Korea; CCMM, Moroccan Coordinated Collections of Microorganisms, Rabat, Morocco; A. Ledeboer, Unilever, Vlaardingen, The Netherlands; ATCC, American Type Culture Collection, Manassas, VA, USA; DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany; NCIMB, National Collections of Industrial, Food and Marine Bacteria, Aberdeen, UK; S. Roos, Swedish University of Agricultural Sciences, Uppsala, Sweden; JCM, Japan Collection of Microorganisms, RIKEN BRC, Japan; J. Björkroth, University of Helsinki, Helsinki, Finland; J. Leisner, Royal Veterinary and Agricultural University, Frederiksberg, Copenhagen, Denmark; CECT, Colección Espanõla de Cultivos Tipo, Valencia, Spain; R. Tenreiro, University of Lisbon, Lisbon, Portugal; L. Dicks, University of Stellenbosch, South Africa. NK, Not known.

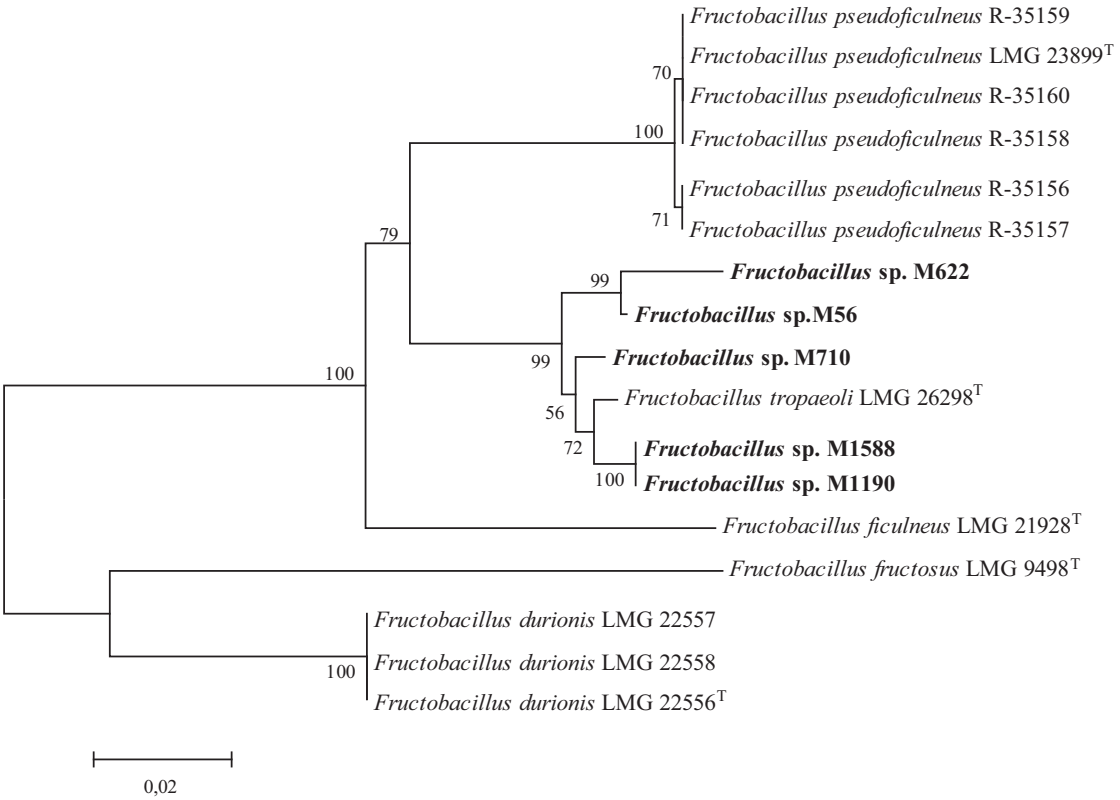


Figure 8: ML tree based on concatenated pheS, rpoA, *and* atpA *gene sequences showing the phylogenetic relationships of strains M56, M622, M710, M1190, and M1588 among other members of the genus* Fructobacillus

Bootstrap values (%) based on 1000 replications are shown at branch points. The substitution model used was the GTR model and the aligned concatenated sequence had a length of 903 bp. Bar, 2% sequence divergence. Accession numbers are given in Table 2.

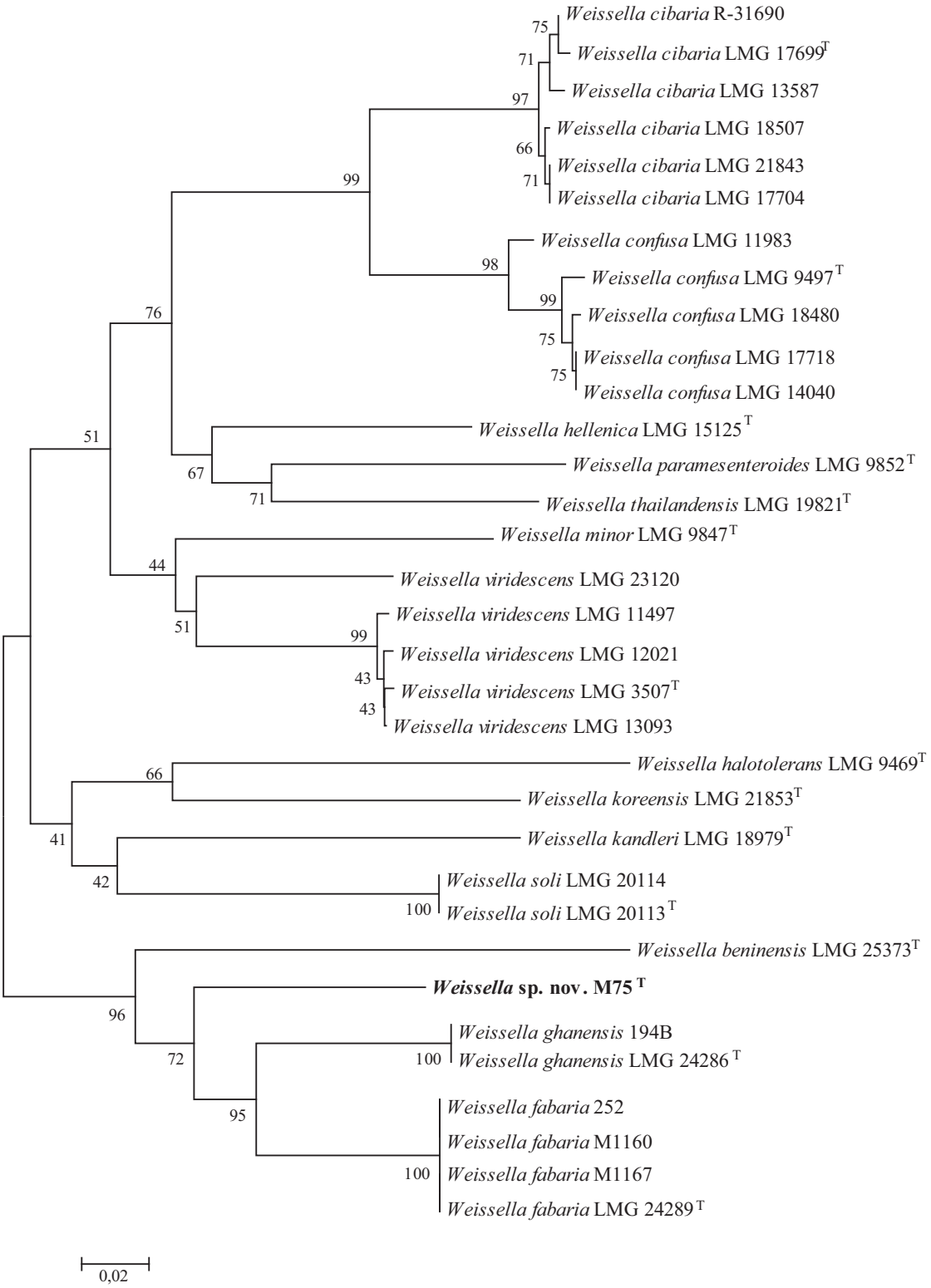


Figure 9: ML tree based on pheS gene sequences showing the phylogenetic relationships of strain M75^T among other members of the genus Weissella

Bootstrap values (%) based on 1000 replications are shown at branch points. The substitution model used was the GTR model and the aligned concatenated sequence had a length of 369 bp. Bar, 2% sequence divergence. Accession numbers are given in Table 2

To confirm these results, DDH experiments between strains M75^T and M1190 and their respective nearest neighbors were performed. Genomic DNA of strain M75^T, *W. ghanensis* LMG 24286^T, *W. fabaria* LMG 24289^T, M1190, and *F. tropaeoli* LMG 26298^T was extracted using the guanidine thiocyanate method described by Pitcher *et al.* (1989). DDHs were performed using the microplate method, with photobiotin for labeling of the DNA (Ezaki *et al.*, 1989), as modified by Goris *et al.* (1998). The hybridization levels of strain M75^T towards *W. ghanensis* LMG 24286^T and *W. fabaria* LMG 24289^T were 44 and 36%, respectively, confirming that it represents a distinct species of the genus *Weissella*. The DDH between strain M1190 and *F. tropaeoli* LMG 26298^T was 82%, which confirmed its MLSA-based identification.

DNA G+C content was determined according to the enzymatic DNA degradation method described previously (Mesbah & Whitman, 1989) using a Waters Breeze high-performance liquid chromatography (HPLC) system and XBridge Shield RP18 column. The solvent used was 0.02 M $NH_4H_2PO_4$ (pH 4.0)/1.5% (v/v) acetonitrile. Non-methylated lambda phage DNA (Sigma) was used as a calibration reference and *Escherichia coli* LMG 2093 DNA was included as a control. The DNA G+C contents of strains M75^T and M1190 were 37 and 45 mol%, respectively, consistent with the G+C contents found previously in the genera *Weissella* (37-47 mol%; (Björkroth *et al.*, 2002; Collins *et al.*, 1993; De Bruyne *et al.*, 2008a; Lee *et al.*, 2002; Magnusson *et al.*, 2002; Padonou *et al.*, 2010; Tanasupawat *et al.*, 2000; Vela *et al.*, 2011) and *Fructobacillus* (42-45 mol%; (Endo & Okada, 2008; Endo *et al.*, 2010).

Preparation of peptidoglycan and analysis of peptidoglycan structure were performed according to published protocols (Schumann, 2011). The total hydrolysate (4 M HCl, 100 °C, 16 h) of the peptidoglycan contained the amino acids lysine, alanine, serine, glycine and glutamic acid in an approximate ratio of 1.0 : 1.7 : 0.9 : 0.2 : 2.6. Because the ratio of glutamic acid to lysine was < 1 and did not fit the expected A3α peptidoglycan type, the peptidoglycan preparation was subjected to hydrofluoric acid treatment in order to remove contaminating polymers linked to the peptidoglycan via phosphodiester bonds before repeating the analysis. However, the molar ratio of the peptidoglycan amino acids changed only slightly after the hydrofluoric acid treatment. The identity of the amino acids was confirmed by gas chromatography coupled to mass spectrometry (GC-MS). The partial hydrolysate (4 M HCl, 100 °C, 45 min) contained, in addition to the amino acids, the peptides L-Ala-D-Glu-L-Lys-L-Ala, D-Ala-L-Lys, and D-Ala-L-Lys-L-Ala. The peptide Ala-Ala could not be found. Dinitrophenylation according to Schleifer (1985) revealed that serine represents the N terminus of the interpeptide bridge.

Though the quantitative amino acid ratio contained too much glutamic acid and too little alanine, the peptidoglycan type A3α L-Lys-L-Ala-L-Ser is concluded from the N terminus of the interpeptide bridge and the 2D thin layer chromatography (TLC) patterns of peptides and amino acids. The high content of glutamic acid as well as the traces of glycine might be caused by residual polymer contamination.

MALDI-TOF MS has shown to be both rapid and accurate for species and subspecies classification of a broad spectrum of bacteria (Sauer & Kliem, 2010). To test whether MALDI-TOF MS analysis is a suitable tool for the identification of the LAB of the present study, mass spectra were generated from all *Weissella* and *Fructobacillus* strains (Table 2) except for *Weissella hellenica* LMG 15125^T, *W. cibaria* R-31690, *W. confusa* LMG 18480, *W. fabaria* 252, and *F. pseudoficulneus* R-35156 and R-35158, from which good-quality spectra could not be obtained. Bacteria were grown under standardized conditions [28 °C, 24 h, de Man-Rogosa-Sharpe (MRS) agar (Oxoid) and an aerobic atmosphere] and subcultured twice prior to analysis. For each strain, a spectral profile was generated from three different generations to improve the robustness of the library of mass spectral fingerprints. Harvesting of cells, extract preparation, measuring, and data analysis were performed as described previously (Ghyselinck *et al.*, 2011). The similarity between the spectra of strains M75^T, M56, M622, M710, M1190, and M1588 and the reference strains was calculated using Pearson's product moment correlation coefficient and clustering was performed using the unweighted pair group method with arithmetic means (UPGMA) clustering algorithm (Figures 4.1 and 4.2, available in Supplementary Material). The MALDI-TOF MS profile of strain M75^T formed a separate cluster that was most similar to the protein profiles of recently described *W. ghanensis* and *W. fabaria* strains originating from Ghanaian cocoa bean fermentations. The remaining cocoa isolates showed high similarity to the recently described *F. tropaeoli*, isolated from a South African flower. For the genus *Fructobacillus*, the intraspecies diversity was consistently smaller than the interspecies divergence towards their nearest neighbors, as was the case for the genus *Weissella*.

Finally, a biochemical analysis of the cocoa isolates was performed to characterize the novel species and to compare the characteristics of the novel *F. tropaeoli* strains (cocoa origin) with those of the type (and currently only known) strain (Chambel *et al.*, 2006; Endo *et al.*, 2010). Cell and colony morphology of strains M56, M622, M710, M1190, and M1588 was verified after growth on MRS agar (Oxoid) supplemented with 1% (w/v) fructose and 24 h of aerobic incubation at 28 °C. For strain M75^T, MRS agar (Oxoid) was used as the basal medium. Conventional bio-

chemical characteristics and enzyme activities were tested as described previously in triplicate, unless stated otherwise (De Bruyne *et al.*, 2007). Growth was tested at 4, 15, 20, 37, and 42°C and in the presence of 6.5, 8, and 10% NaCl (w/v). The production of gas from glucose was determined using inverted Durham tubes. The API 50 CHL *Lactobacillus* identification system (bioMérieux) proved useful for determining carbohydrate fermentation profiles. The production of D- and L-lactate from glucose was determined enzymatically (B-Biopharm). Unlike the *F. tropaeoli* isolates, strain M75^T was able to produce dextran in MRS agar (Oxoid) in which glucose had been replaced with 5% sucrose. The physiological and biochemical characteristics of the *F. tropaeoli* isolates were identical to those of *F. tropaeoli* LMG 26298^T, with the only difference that the isolates were also able to ferment sucrose, in addition to D-glucose, D-fructose, and D-mannitol. The results for strain M75^T are given in the species description. Characteristics that differentiate strain M75^T from other species of the genus *Weissella* are summarized in Table 3.

4.1.3 Conclusion

In conclusion, the results of the present study demonstrate that strains M56, M622, M710, M1190, and M1588 belong to *F. tropaeoli*. These results and those of Papalexandratou (2011a; 2011c; 2011d) demonstrate that *F. tropaeoli* participates in the fermentation process of cocoa beans in Brazil, Ecuador, and Malaysia, which are major cocoa-producing countries. Additionally, the results of the present study demonstrate that strain M75^T represents a novel species of the genus *Weissella* that is closely related to *W. fabaria*, *W. ghanensis*, and *W. beninensis* (96.9, 94.9, and 94.2% 16S rRNA gene sequence similarity towards the respective type strains), but which can be distinguished from these and other species of the genus *Weissella* by DDH, *pheS* gene sequence analysis, MALDI-TOF MS-based analysis, and biochemical characteristics. Based on these results, we propose to classify strain M75^T as the type strain of the novel species *Weissella fabalis* sp. nov.

*Table 3: Differential characteristics of strain M75^T (*W. fabalis *sp. nov.) and other species of the genus* Weissella

Characteristic	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Acid from:															
Arabinose	-	-	-	-	-	-	+	-	+	+	-	+	+	d	-
Cellobiose	+	+	+	-	+	-	-	-	-	+	+	-	-	d	d
Fructose	+	+	+	+	+	+	-	+	ND	+	+	+	+	+	+
Galactose	-	-	-	-	-	-	-	+	-	-	+	+	-	+	+
Maltose	+	-	+	+	+	+	+	-	-	+	+	+	+	+	+
Melibiose	-	-	-	-	-	-	+	-	-	-	-	+	-	+	+
Raffinose	-	-	-	-	-	-	+	-	-	-	-	+	-	d	+
Ribose	-	-	-	+	+	-	+	+	+	-	+	+	-	d	d
Salicin	-	-	+	-	-	-	+	-	ND	+	+	-	-	-	d
Sucrose	-	-	d	-	+	d	+	-	-	+	+	d	+	+	+
Trehalose	+	+	+	-	+	d	+	-	-	-	-	d	+	+	d
Xylose	-	-	-	-	-	-	+	-	+	+	+	-	-	d	-
Hydrolysis from esculin	+	+	+	-	+	-	+	-	-	+	+	-	ND	d	+
NH_3 from arginine	+	+	+	+	+	-	+	+	+	+	+	-	-	-	+
Dextran formation	+	+	+	ND	-	ND	-	+	+	+	+	-	-	-	ND
Lactic acid configuration	D	DL	DL	DL	DL	DL	D	D	D	DL	D	D	D	D	D
Growth at/in:															
15°C	+	+	+	ND	ND	ND	+	ND	+	+	ND	ND	ND	ND	+
37°C	+	+	+	ND	-	ND	-	+	+	+	+	-	-	-	ND
42°C	+	+	+	ND	ND	ND	+	ND	+	+	ND	+	ND	ND	ND
6.5% NaCl	-	-	-	ND	ND	ND	-	ND	-	+	ND	-	ND	ND	-
8.0% NaCl	-	-	-	ND	ND	ND	ND	ND	ND	+	ND	ND	ND	ND	ND
10.0% NaCl	-	-	-	ND	ND	ND	ND	ND	-	-	ND	ND	ND	ND	ND
DNA G+C content (mol%)	37	38	40	44	44	41-44	43	39	37	44-45	45-47	38-41	39-40	37-38	37

Species: 1, *W. fabalis* sp. nov.; 2, *W. fabaria*; 3, *W. ghanensis*; 4, *W. halotolerans*; 5, *W. minor*; 6, *W. viridescens*; 7, *W. soli*; 8, *W. kandleri*; 9, *W. koreensis*; 10, *W. cibaria*; 11, *W. confusa*; 12, *W. thailandensis*; 13, *W. hellenica*; 14, *W. paramesenteroides*; 15, *W. beninensis*. +, 90% or more strains positive; 2, 90% or more strains negative; d, 11-89% of strains positive; ND, no data available. Data partially adapted from (Collins *et al.*, 1993; Tanasupawat *et al.*, 2000; De Bruyne *et al.*, 2008a,0; Padonou *et al.*, 2010).

Description of Weissella fabalis *sp. nov.*

Weissella fabalis (fa.ba'lis. L. fem. adj. *fabalis* belonging to beans. Cells are Gram-stain-positive, catalase-negative, facultatively anaerobic, and non-motile. Cells are coccoid, approximately 1.0 μm wide and 1.5 μm long, and occur singly, in pairs or in short chains. Colonies grown for 2 days on MRS agar at 28°C are approximately 1 mm in diameter, beige, opaque, smooth, and circular, with a low-convex elevation. Gas is produced from glucose, indicating the heterofermentative character of the type strain. Produces D-lactic acid. Growth occurs at 15-37°C and in the presence of 5-6% NaCl but not in the presence of 7-8% NaCl. Arginine is hydrolyzed. Acid is produced from glucose, fructose, mannose, N-acetylglucosamine, aesculin, cellobiose, maltose, trehalose, and gentiobiose. Acid is not produced from glycerol, erythritol, D- or L-arabinose, ribose, D- or L-xylose, adonitol, methyl β-D-xylopyranoside, galactose, sorbose, rhamnose, dulcitol, inositol, mannitol, sorbitol, methyl α-D-mannopyranoside, methyl α-D-glucopyranoside, amygdalin, arbutin, salicin, lactose, melibiose, sucrose, inulin, melezitose, raffinose, starch, glycogen, xylitol, turanose, D-lyxose, D-tagatose, D- or L-fucose, D- or L-arabitol, potassium gluconate, potassium 2-ketogluonate, or potassium 5-ketogluconate. The DNA G+C content of the type strain is 37 mol%.

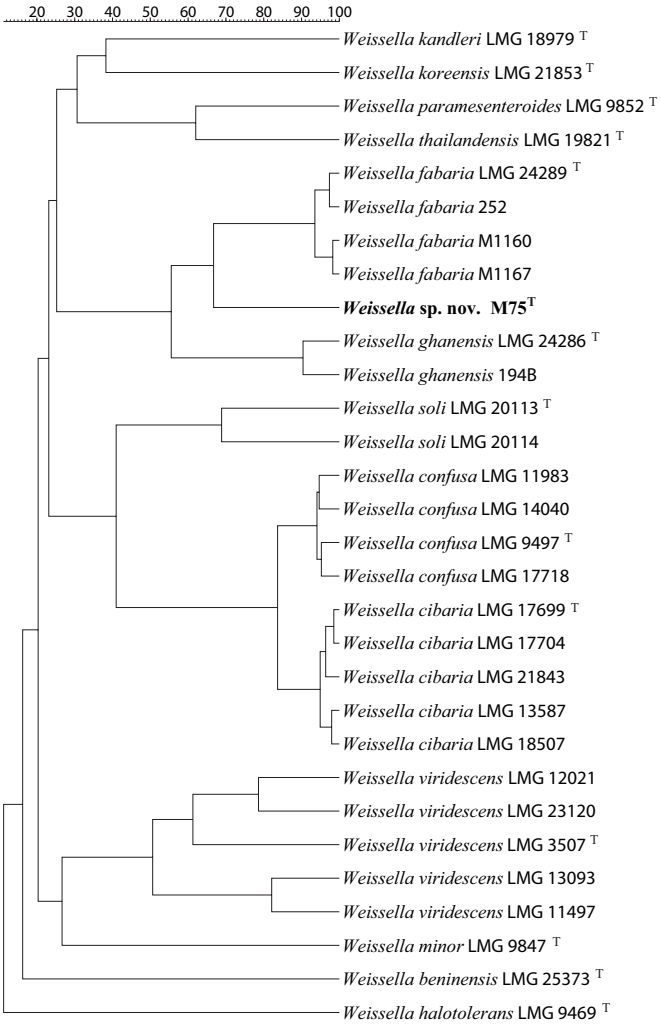
The type strain, M75^T (=LMG 26217^T =CCUG 61217^T), was isolated from a Brazilian cocoa bean box fermentation carried out in Ilhéus, Bahia, Brazil, in 2007.

4.1.4 Acknowledgements

We acknowledge financial support from the Fund for Scientific Research (FWO)-Flanders and thank Anneleen Wieme and Freek Spitaels for their contributions to the development of the MALDI-TOF MS protocol. Financial support from the Research Council of the Vrije Universiteit Brussel and Barry Callebaut N.V. is also acknowledged.

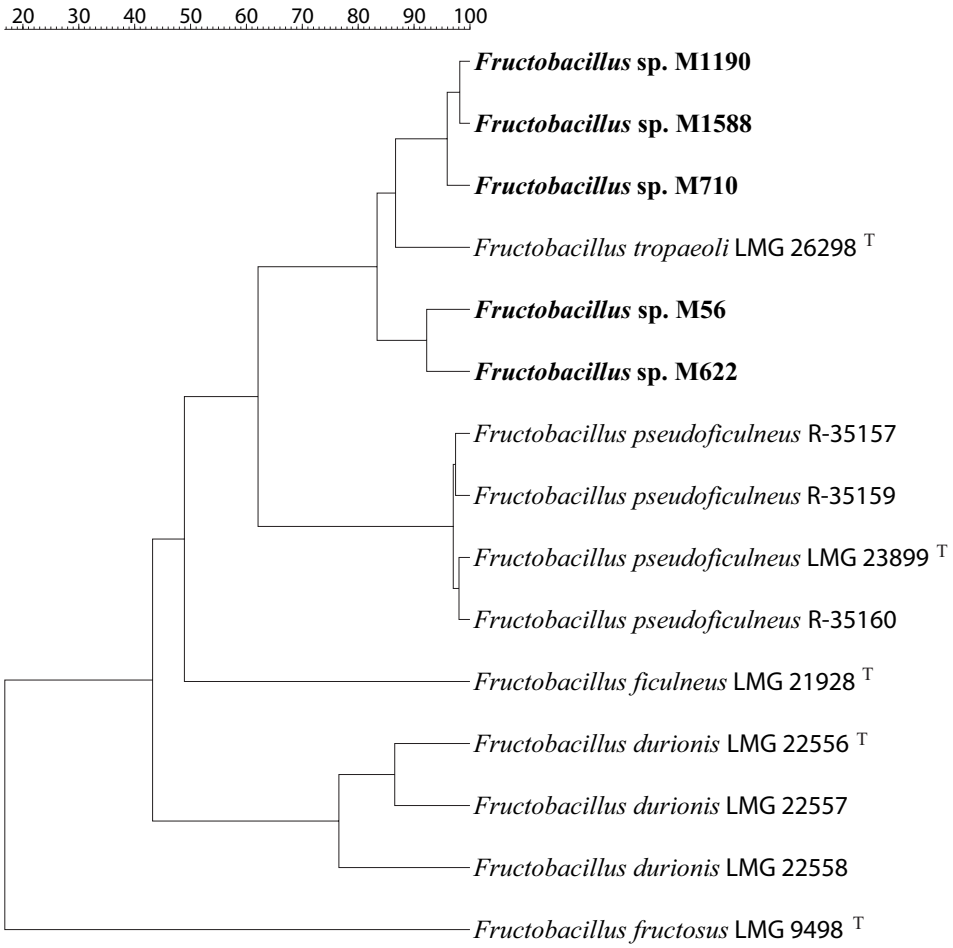
4.1.5 Supplementary Material

The GenBank/EMBL/DDBJ (DNA Data Bank of Japan) accession numbers for the 16S rRNA and *pheS* gene sequences of strain M75^T are HE576795 and HE576796. Accession numbers of other sequences obtained in this study are detailed in Table 2.



Supplementary Material 4.1: MALDI-TOF MS dendrogram constructed using the Pearson product moment correlation coefficient and the UPGMA clustering algorithm, showing the similarity of strain M75^T with other strains of the genus Weissella.

All strains of Table 2 are incorporated in the dendrogram, with the exception of LMG 15125^T, R-31690, and LMG 18480.



Supplementary Material 4.2: MALDI-TOF MS dendrogram constructed using the Pearson product moment correlation coefficient and the UPGMA clustering algorithm, showing the similarity of strains M56, M622, M710, M1190, and M1588 with other strains of the genus Fructobacillus.

All strains of Table 2 are incorporated in the dendrogram, with the exception of R-35156 and R-35158.

4.2 *Leuconostoc rapi* sp. nov., Isolated from Sous-Vide Cooked Rutabaga

U. Lyhs, I. Snauwaert, S. Pihlajaviita, L. De Vuyst, P. Vandamme

Submitted to: International Journal of Systematic and Evolutionary Microbiology, 2014

Author contributions: UL wrote the manuscript. IS performed the experiments. SP did the isolation of the strains. PV and IS designed the experiment. PV and LDV proofread the manuscript.

A Gram-stain-positive, ovoid LAB, strain LMG 27676^T, was isolated from a spoiled sous-vide cooked rutabaga. 16S rRNA gene sequence analysis indicated that the novel strain belongs to the genus *Leuconostoc*, with *Leuconostoc kimchi* and *Leuconostoc miyukkimchii* as nearest neighbors (99.1 and 98.8% 16S rRNA gene sequence similarity towards the type strain, respectively). Phylogenetic analysis of the 16S rRNA gene, MLSA of the *pheS*, *rpoA*, and *atpA* genes, and biochemical and genotypic characterization allowed to differentiate strain LMG 27676^T from all established *Leuconostoc* species. Strain LMG 27676^T (= DSMZ 27776^T) therefore represents a new species, for which the name *Leuconostoc rapi* sp. nov. is proposed.

4.2.1 Introduction

The genus *Leuconostoc*, which at the time of writing comprises 12 species, belongs phylogenetically to the phylum *Firmicutes*, class Bacilli, order *Lactobacillales*. *Leuconostoc* cells are Gram-stain-positive, small, regular, ovoid cocci, occurring in pairs or short chains. The cells are non-motile and do not form spores. *Leuconostoc* species are considered psychrotrophic mesophiles with optimal growth at 14-30°C. The temperature limits for growth vary among species and strains, ranging from 1-10°C to 30-40°C (Säde, 2011). In 2008, *L. fructosum*, *L. durionis*, *L. ficulneum*, and *L. pseudoficulneum* were reclassified on the basis of phylogenetic, physiological, and morphological differences into the genus *Fructobacillus* gen. nov. (Endo & Okada, 2008).

Species of the genus *Leuconostoc* are associated with the fermentation of vegetables (Di Cagno *et al.*, 2013). Latest described *Leuconostoc* species originating

from fermented vegetables include *L. inhae* and *L. kimchii* from kimchi (Kim *et al.*, 2000) and *L. miyukkimchii* from fermented miyukkimchii made of a brown algae (*Undaria pinnatifida*) and is a regional kimchi in Korea (Lee *et al.*, 2012). Other novel species include *L. palmae* from palm wine, an alcoholic beverage produced from the sap of various palm tree species (Ehrmann *et al.*, 2009), and *L. holzapfelii*, from Ethiopian coffee fermentation (De Bruyne *et al.*, 2007). The occurrence of leuconostocs including *L. gelidum*, *L. gasicomitatum* (both of which were recently reclassified as *L. gelidum* subsp. *gelidum* and *L. gelidum* subsp. *gasicomitatum*, respectively (Rahkila *et al.*, 2014)), *L. citreum*, and *L. mesenteroides*, in other vegetables products has also been reported (García-Gimeno & Zurera-Cosano, 1997; Lyhs *et al.*, 2004; Vihavainen *et al.*, 2008).

Strain LMG 27676^{T} was isolated from packaged sous-vide cooked rutabaga (*Brassica napobrassica*) during an investigation of the spoilage microbiota in sous-vide cooked vegetables in Finland. For microbiological analyses, 10 g rutabaga sample was aseptically weighed into 90 ml of buffered peptone water (BPW) (LabM, Lancashire, UK) in a sterile plastic bag and then blended for 60 s using a Dilumat automatic system (bioMérieux). Tenfold serial dilutions were used for microbiological analyses. The number of LAB was determined on MRS agar (Oxoid) supplemented with sorbic acid (pH 6.4). All plates were incubated under anaerobic conditions at 20°C for 7 days. Three of the isolates obtained, LMG 27676^{T}, R-50028, and R-50030, exhibited identical RAPD fingerprints as determined using the primers RAPD-270 (5'-TGCGCGCGGG-3') and RAPD-272 (5'-AGCGGGCCAA-3') (Mahenthiralingam *et al.*, 1996) (data not shown).

4.2.2 Methods, Results, and Discussion

Nearly complete 16S rRNA gene sequence analysis was performed as described by Vancanneyt *et al.* (2004) for strain LMG 27676^{T}. PCR products were purified using a Nucleofast 96 PCR clean up membrane system (Machery-Nagel, Düren, Germany). The sequencing primers were those listed by Coenye *et al.* (1999) and the fragments obtained were cleaned with a BigDye XTerminator Purification kit (Applied Biosystems, Life Technologies, Carlsbad, CA, USA), according to the manufacturer's instructions. The mothur software package and the corresponding SILVA reference alignment were used to align the 16S rRNA gene sequences of strain LMG 27676^{T} (1507 bp) with sequences of type strains of all established species of the genus *Leuconostoc*. These aligned sequences were imported into MEGA version 5.0 (Tamura *et al.*, 2011) and analyzed using the neighbor-joining,

ML, and maximum-parsimony methods. The statistical reliability of tree topologies was evaluated by bootstrapping analysis based on 1000 replicates. The neighbor-joining and the maximum-parsimony trees revealed topologies similar to those obtained in the ML tree (Figure 10). 16S rRNA gene sequence similarities were calculated using MEGA version 5.0 (Tamura *et al.*, 2011) and the closest relatives of strain LMG 27676^T were *L. kimchii* IH25^T (99.1%) and *L. miyukkimchii* M2^T (98.8%). Lower sequence similarities (< 98.6%) were found with other members of the genus *Leuconostoc*.

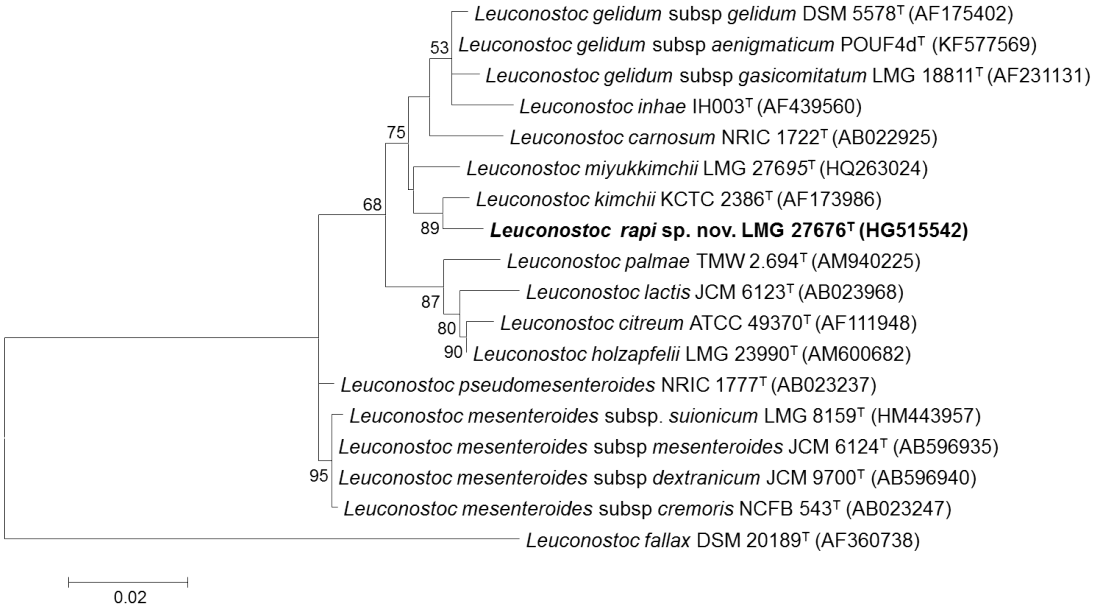


Figure 10: ML tree of the 16S rRNA gene sequence of strain LMG 27676^T with all type strains of the genus Leuconostoc

The evolutionary history was inferred by using the ML method based on the Kimura 2-paramter model. The tree with the highest log likelihood is shown. There were a total of 1379 positions in the final dataset.

Sequence analysis of the housekeeping genes encoding *pheS*, *rpoA*, and *atpA* correlates with a species delineation as determined by DDH in the genus *Leuconostoc* (De Bruyne *et al.*, 2007). Therefore, *pheS*, *rpoA*, and *atpA* gene sequences of strains *Leuconostoc rapi* sp. nov. LMG 27676^T, *Leuconostoc kimchii* KCTC 2386^T, and *Leuconostoc miyukkimchii* LMG 27695^T, for which no *pheS*, *rpoA*, or *atpA* sequences were available, were determined as described previously (De Bruyne *et al.*, 2007), except that sequencing reactions were purified using the BigDye xTerminator purification kit as described above. The *pheS*, *rpoA*, and *atpA* gene

sequences of the closest relatives (*L. carnosum* LMG 23898^T, *L. inhae* LMG 22919^T, *L. gelidum* subsp. *gasicomitatum* LMG 18811^T, and *L. gelidum* subsp. *gelidum* LMG 18297^T, and *L. gelidum* subsp. *aenigmaticum* LMG 27840^T) of strain LMG 27676^T were available from previous studies (De Bruyne *et al.*, 2007; Rahkila *et al.*, 2014). The software package MEGA version 5.0 (Tamura *et al.*, 2011) was used to align the translated concatenated gene sequences and to analyze the nucleotide sequences as mentioned above. The maximum-parsimony and neighbor-joining trees (not shown) revealed topologies similar to those obtained in a phylogenetic tree reconstructed using the ML approach (Figure 11) for both analyses. The concatenated *pheS*, *rpoA*, and *atpA* gene sequences of strain LMG 27676^T revealed high similarity to the concatenated sequence of *L. kimchii* KCTC 2386^T (94.2%). Lower concatenated sequence similarities ($<$ 88.1%) were found towards other species of the genus *Leuconostoc* with validly published names. The *pheS*, *rpoA*, and *atpA* gene sequences of strain LMG 27676^T revealed high similarities to the sequences of *L. kimchii* KCTC 2386^T (91.3, 96.4, and 94.4% similarity, respectively). Lower gene sequence similarities were found ($<$ 84.5, 89 and 91.5%, respectively) towards other species of the genus *Leuconostoc* with validly published names. According to De Bruyne and coworkers (2007), species now classified in the genera *Leuconostoc* and *Fructobacillus* are delineated above 93, 98, and 98% *pheS*, *rpoA*, and *atpA* gene sequence similarity, respectively. Therefore, these results suggest that strain LMG 27676^T represents a novel species of the genus *Leuconostoc*. To confirm this, a DDH experiment between strain LMG 27676^T and *L. kimchii* strains KCTC 2386^T and LMG 23786 was performed. Genomic DNA was extracted using the guanidine thiocyanate method described by Pitcher *et al.* (1989). The DDH experiment was performed using the microplate method, with photobiotin for labeling of the DNA (Ezaki *et al.*, 1989), as modified by Goris *et al.* (1998). The hybridization levels of strain LMG 27676^T towards *L. kimchii* KCTC 2386^T was 51% (the reciprocal hybridization values were 53 and 49%); the hybridization level between *L. kimchii* KCTC 2386^T and *L. kimchii* LMG 23786 was 85% (the reciprocal hybridization values were 72 and 97%); and the hybridization level between strain LMG 27676^T and *L. kimchii* LMG 23786 was 64% (the reciprocal hybridization values were 70 and 57%). The hybridization values between strain LMG 27676^T and both *L. kimchi* strains were clearly below the species delineation threshold.

DNA G+C content was determined according to the enzymatic DNA degradation method described previously (Mesbah & Whitman, 1989), using a Waters Breeze HPLC system and XBridge Shield RP18 column. The solvent used was 0.02 M $NH_4H_2PO_4$ (pH 4.0)/1.5% (v/v) acetonitrile. Non-methylated lambda phage DNA (Sigma-Aldrich, St. Louis, MO, USA) was used as a calibration reference and *Escherichia coli* LMG 2093 DNA was included as a control. The DNA G+C content

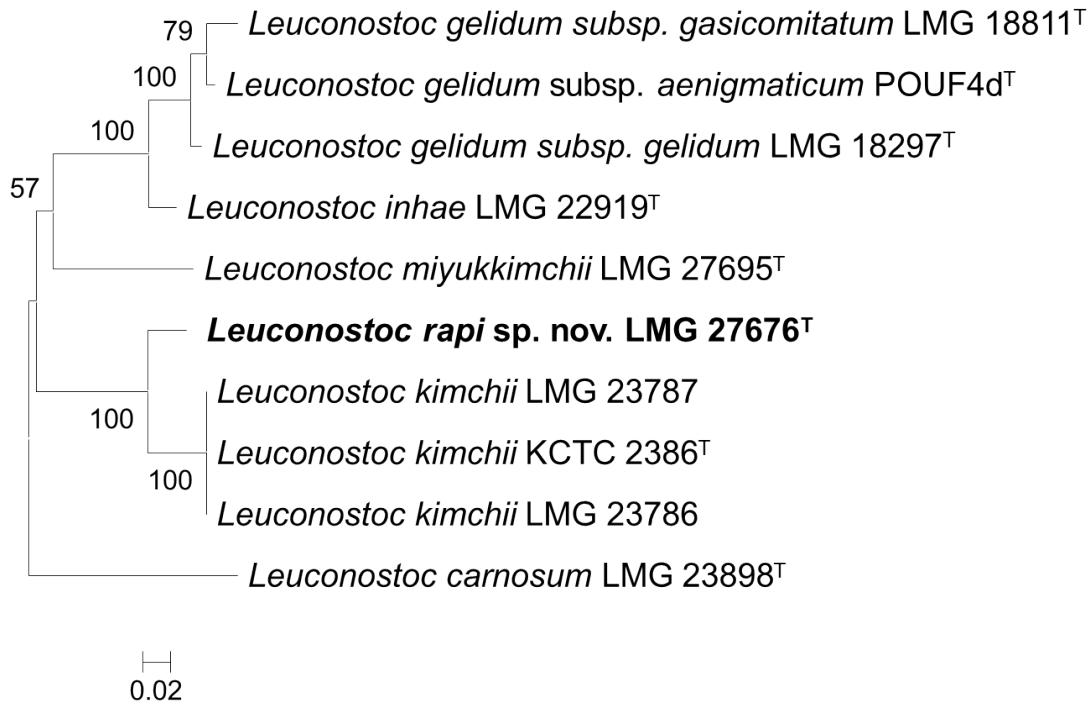


Figure 11: ML tree of the concatenated pheS, rpoA, *and* atpA *gene sequences of the closest relatives of strain LMG 27676^T*

The evolutionary history was inferred by using the ML method based on the Tamura 3-parameter model. The tree with the highest log likelihood is shown. There were a total of 1482 positions in the final dataset. The type strain of *L. kimchii* was not available and was consequently replaced by strain LMG 23787

of strain LMG 27676^T was 38 mol% which is consistent with the G+C contents found previously in the genus *Leuconostoc* (38-44 mol%; (De Bruyne *et al.*, 2007).

Preparation of peptidoglycan and analysis of the peptidoglycan structure were performed according to published protocols (Schumann, 2011). The total hydrolysate (4 N HCl, 100°C, 16 h) contained the amino acids alanine, glutamic acid and lysine, and analysis of partial hydrolysates (4 N HCl, 100°C, 45 min) revealed the presence of the peptidoglycan type A3α L-Lys-Ala$_2$ or type A11.5 according to `www.peptidoglycan-types.info`.

Finally, a biochemical analysis of LMG 27676^T, KCTC 2386^T, and LMG 23786 was performed to characterize the novel species. Cell and colony morphology of

strain LMG 27676^T was verified after growth on MRS agar (Oxoid) and 96 h of aerobic incubation at 28°C. Conventional biochemical characteristics and enzyme activities were tested as described previously in triplicate, unless stated otherwise (De Bruyne *et al.*, 2007). Growth was tested at 4, 20, 30, 37, and 45°C and in the presence of 0, 2, 4, 6, 8, and 10% NaCl (w/v). The production of gas from glucose was determined using inverted Durham tubes and carbohydrate fermentation profiles were determined using the API 50 CHL *Lactobacillus* identification system (bioMérieux). Furthermore, tests for the procution of ammonia from arginine were performed as described previously (De Bruyne *et al.*, 2007). To test for the production of D- and L-lactate from glucose, cells were grown in MRS broth with glucose as the sole energy source. Proteins were removed by adding an equal volume of acetonitrile (Sigma-Aldrich, St. Louis, MO, USA); samples were subsequently microcentrifuged (14000 rpm for 8 min), filtered (0.2 μm, Sartorius AG, Göttingen, Germany) and transferred to an appropriate vessel prior to injection. The amount of D- and L-lactic acid in the supernatant was determined using HPLC with ultra violet (UV) detection. To this end, a Waters chromatograph (Waters Corporation, Milford, MA, USA) was used, equipped with a 486 UV detector, a 600S controller, a 717Plus autosampler, and a Shodex-Column Orpak CRX-853; 50 x 8 mm (Showa Denko KK, Tokyo, Japan). The mobile phase, at a flow rate of 1 ml min^{-1}, consisted of a 1 mM $CuSO_4$ solution in ultrapure water with 20% acetonitrile (Sigma-Aldrich). D- and L-lactic acid were eluted isocratically and were detected by measuring the absorption at 253 nm. The results for strain LMG 27676^T are given in the species description below. Characteristics that differentiate strain LMG 27676^T from its nearest phylogenetic neighbors are summarized in Table 4. The biochemical characteristics of KCTC 2386^T and LMG 23786 were identical with the following exceptions: only KCTC 2386^T produces acid from L-arabinose; LMG 23786 produces acid from arbutin, which is not the case for KCTC 2386^T.

Table 4: Differential phenotypic characteristics of strain LMG 27676^T and closest related species of the genus Leuconostoc

Characteristic	1	2	3
Source	Rutabaga	Kimchi	Brown algae kimchi
Morphology	Ovoid	Coccus	Ovoid
NaCl (%) range for growth	0-6	0-7	0-6
Growth at 4°C	-	+	-
Growth at 37°C	-	+	+
Acid from:			
L-Arabinose	+	+	-
Arbutin	-	-[a]	+
Starch	-	-	+
D-xylose	-	-	+
DNA G+C content (mol%)	38	37	42.5

1: *L. rapi* sp. nov LMG 27676^T, 2: *L. kimchii* KCTC 2386^{TT} and 3: *L. miyukkimchii* LMG 27965^T.
[a]: This biochemical characteristic differs from what was previously reported (Kim *et al.*, 2000).

4.2.3 Conclusion

In conclusion, the results of the present study demonstrate that strain LMG 27676^T represents a taxon that can be differentiated from the present *Leuconostoc* species by a range of genotypic methods and by several biochemical characteristics including the absence of growth at 37°C. Although the genotypic divergence in terms of DDH towards its nearest neighbor *L. kimchii* KCTC 2386^T yielded a value that was close to the threshold level for species delineation, the *pheS*, *rpoA*, and *atpA* gene sequences clearly differentiated both taxa genotypically. We therefore classify the taxon represented by strain LMG 27676^T as a novel *Leuconostoc* species, for which we propose the name *Leuconostoc rapi* sp. nov.

Description of Leuconostoc rapi *sp. nov.*

Leuconostoc rapi (ra'pi. L. gen. n. rapi of a turnip, referring to rutabaga, a turnip-like vegetable). Cells are Gram-stain-positive, catalase-negative, facultatively anaerobic, and non-motile. Cells are ovoid and occur singly, in pairs or in short chains. Colonies grown for 3 days on MRS agar at 28°C are approximately 1 mm in diameter, white/creamy, convex, and circular with smooth margins. Gas is produced from glucose, indicating the heterofermentative character of the type

strain. Produces L- and D-lactic acid in a ratio of 2:7. Growth occurs at 20-30°C and in the presence of 0-6% NaCl but not the in the presence of 8-10% NaCl. Arginine is not hydrolyzed. Acid is produced from glucose, fructose, mannose, L-arabinose, D-ribose, mannitol, methyl-α-D-glucopyranoside, N-acetylglucosamine, amygdalin, esculin ferric citrate, salicin, D-cellobiose, sucrose, threhalose, gentiobiose, turanose, potassium gluconate, and potassium 2-ketogluconate. Acid is not produced from glycerol, erythritol, D-arabinose, D- and L-xylose, D-adonitol, methyl-β-D-xylopyronoside, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-sorbitol, methyl-α-D-mannopyranoside, arbutin, D-maltose, D-lactose, D-melibiose, inulin, D-melezitose, D-raffinose, starch, glycogen, xylitol, D-lyxose, D-tagatose, D- and L-fucose, and D- and L-arabitol. The DNA G+C content of the type strain is 38 mol% and its peptidoglycan type is A3α L-Lys-Ala_2. The type strain, LMG 27676^T (= DSMZ 27776^T) was isolated from a spoiled sous-vide cooked rutabaga produced by a Finnish manufacturer.

4.2.4 Acknowledgements

The authors would like to thank Kauhajoki Food Laboratory and Pauliina Isohanni (Ruralia Institute, Seinäjoki Unit, University of Helsinki) for their technical assistance. Prof. Anu Hopia (Functional Foods Forum, University of Turku) and Prof. Per Saris (Department of Applied Chemistry and Microbiology, Faculty of Agriculture and Forestry, University of Helsinki) are acknowledged. The strain was isolated in a research project 'Novel food technologies - Special focus on quality and sustainable processing' that was financed by European Regional Development Fund (ERDF) via Tekes (The Finnish Funding Agency for Innovation), University Foundation of South Ostrobothnia and participating companies.

4.2.5 Supplementary Material

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA, *pheS*, *rpoA*, and *atpA* gene of strain LMG 27676^T are HG515542, HG515543, HG515544, and HG515545. The accession numbers for the *pheS*, *rpoA*, and *atpA* gene of strain LMG 27695^T (*L. miyukkimchii*) and KCTC 2386^T (*L. kimchii*)are HG515546-HG515548 and LN650690-LN650692, respectively.

4.3 *Carnobacterium iners* sp. nov., a Psychrophilic, Lactic Acid-Producing Bacterium from the Littoral Zone of an Antarctic Pond.

I. Snauwaert, B. Hoste, K. De Bruyne, K. Peeters, L. De Vuyst, A. Willems, P. Vandamme

Redrafted from: International Journal of Systematic and Evolutionary Microbiology, 2013 (63), 1370-75

Author contributions: IS performed the experiments and wrote the manuscript. BH and KDB designed the primers for the MLSA scheme construction. KP did the isolation and dereplication of the strains. PV and IS designed the experiment. AW, LDV, and PV proofread the manuscript.

Two lactic acid-producing, Gram-stain-positive rods were isolated from a microbial mat actively growing in the littoral zone of an Antarctic lake (Forlidas Pond) in the Pensacola mountains and studied using a polyphasic taxonomic approach. The isolates were examined by phylogenetic analysis of the 16S rRNA gene, MLSA of *pheS*, *rpoA*, and *atpA*, and biochemical and genotypic characteristics. One strain, designated LMG 26641, belonged to *Carnobacterium alterfunditum* and the other strain, designated LMG 26642^T, could be assigned to a novel species, with *Carnobacterium funditum* DSM 5970^T as its closest phylogenetic neighbor (99.2% 16S rRNA gene sequence similarity). *Carnobacterium iners* sp. nov. could be distinguished biochemically from other members of the genus *Carnobacterium* by the lack of acid production from most carbohydrates. DNA-DNA relatedness confirmed that strain LMG 26642^T represented a novel species, for which we propose the name *Carnobacterium iners* sp. nov. (type strain is LMG 26642^T = CCUG 62000^T).

4.3.1 Introduction

The genus *Carnobacterium* was created to accommodate heterofermentative, facultatively anaerobic, psychrotolerant, rod-shaped LAB that produce L-lactic acid from glucose. At the time of writing, the genus *Carnobacterium* comprises 10 species with validly published names. Most of them were isolated from food of animal origin and/or living fish, except for *Carnobacterium alterfunditum*, *Carnobacterium funditum* and *Carnobacterium pleistocenium* which were isolated from cold environments with low nutrient contents, such as Antarctic ice lakes and permafrost ice (Franzmann *et al.*, 1991; Pikuta *et al.*, 2005).

Recently, Peeters *et al.* (2011) examined the heterotrophic bacterial diversity in a microbial mat sample originating from the littoral zone of a continental Antarctic lake (Forlidas Pond) in the Pensacola mountains. A large number of bacteria were isolated and characterized through rep-PCR fingerprinting and phylogenetic analysis of partial 16S rRNA gene sequences. Thirty LAB belonging to two main rep-clusters were isolated on marine agar 2216 (BD Difco) at 4-20°C in an aerobic or anaerobic atmosphere. Representatives of rep-cluster I (LMG 26642^T, R-36994, R-37000, and LMG 26896) and rep-cluster II (LMG 26641 and LMG 26897) were grown aerobically on trypticase soy broth with yeast extract salt medium [TSBY salt; 3% trypticase soy broth (Oxoid), 0.3% yeast extract (Oxoid), 1.3% NaCl (Merck), 0.034% KCl (UCB), 0.4% $MgCl_2.6H_2O$ (Sigma-Aldrich), 0.345% $MgSO_4.7H_2O$ (UCB), 0.025% NH_4Cl (Merck), 0.014% $CaCl_2.2H_2O$ (Merck), 2% agar no. 2 (LabM), pH 7.2] at 4°C for 10 days and 20°C for 4 days, respectively. The clonality of these isolates was investigated by RAPD fingerprinting using primers RAPD-270 (5'-TGCGCGCGGG-3') and RAPD-272 (5'-AGCGGGCCAA-3'), as described by Mahenthiralingam *et al.* (1996). The fingerprints indicated that isolates LMG 26642^T, LMG 26896, R-37000, and R-36994 were genetically identical, as was the case for LMG 26641 and LMG 26897 (data not shown).

4.3.2 Methods, Results, and Discussion

Nearly complete 16S rRNA gene sequence analysis was performed as described by Vancanneyt *et al.* (2004) for strains LMG 26642^T and LMG 26641. PCR products were purified using a Nucleofast 96 PCR clean up membrane system (Macherey-Nagel, Germany). The sequencing primers were those listed by Coenye *et al.* (1999) and the obtained fragments were cleaned with a BigDye xTerminator Purification kit (Applied Biosystems, USA), according to the manufacturer's in-

structions. ARB (Ludwig *et al.*, 2004) and the corresponding SILVA SSURef 102 database (Pruesse *et al.*, 2007) were used to align the 16S rRNA gene sequences of strains LMG 26642^T (1512 bp) and LMG 26641 (1433 bp) with sequences of type strains of all established species of the genus *Carnobacterium*. These aligned sequences were imported into MEGA version 5.0 (Tamura *et al.*, 2011) and analyzed using the neighbor-joining, ML and maximum-parsimony methods. The statistical reliability of tree topologies was evaluated by bootstrapping analysis based on 1000 replicates. The neighbor-joining and the maximum-parsimony trees revealed topologies similar to that obtained in the ML tree (Figure 12). 16S rRNA gene sequence similarities were calculated using ARB and the closest relatives of strain LMG 26642^T were *C. funditum* DSM 5970^T (99.2%) and *Carnobacterium mobile* DSM 4848^T (97.0%). Strain LMG 26641 showed 100 and 99.8% 16S rRNA gene sequence similarity with *C. pleistocenium* FTR1^T and *C. alterfunditum* ACAM 311, respectively. Lower sequence similarities (< 99.2%) were found with other members of the genus *Carnobacterium*.

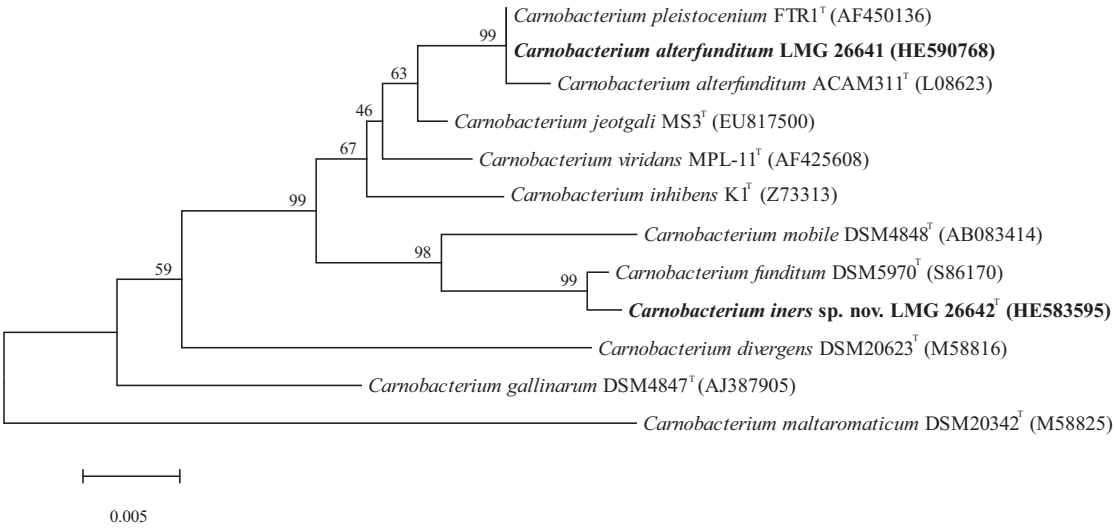


Figure 12: ML tree based on 16S rRNA gene sequences showing the phylogenetic relationship of strains LMG 26642^T and LMG 26641 with type strains of species of the genus Carnobacterium

Bootstrap values (> 50%) based on 1000 replications are shown at branch nodes. Bar, 0.5% sequence divergence. The substitution model used was the GTR model and the aligned sequence had a length of 1396 bp.

Because of the failure of 16S rRNA gene sequence similarity to reflect the phylogeny of closely related LAB, MLSA was performed to obtain a higher taxonomic

resolution. MLSA of the housekeeping genes encoding *pheS*, *rpoA*, and *atpA* has been successfully applied for the identification of different LAB (De Bruyne *et al.*, 2007; Naser *et al.*, 2005a). The primer combinations used for the amplification and sequencing of *pheS*, *rpoA*, and *atpA* are listed Table 5. Primer combinations *pheS*-24/25-F/*pheS*-26/27-R, *rpoA*-24-F/*rpoA*-25/26-R, and *atpA*-29-F/*atpA*-35-R amplified the target genes of all strains except for *atpA* of *C. alterfunditum*, *C. pleistocenium*, and *C. jeotgali*; for these, an alternative primer combination (*atpA*-30-F/*atpA*-33/34-R) was used. Amplification conditions and sequencing reactions were performed as described by Naser *et al.* (2005a) with the following modifications: (i) the annealing temperature was set at 50°C but in a few cases this value was raised to 55°C to avoid aspecific amplicons and (ii) sequencing reactions were purified using a BigDye xTerminator Purification kit. To assess the inter- and intraspecies variation, multiple strains per species were included if available; in total, 23 strains were studied. The strains examined and their depositors, sources and accession numbers are listed in Table 6. SeaView version 4 was used to concatenate the gene sequences of *pheS*, *rpoA*, and *atpA* (Gouy *et al.*, 2010). MEGA version 5.0 (Tamura *et al.*, 2011) was used to align the translated concatenated gene sequences and analyze the nucleotide sequences. The neighbor-joining and maximum-parsimony tree revealed topologies similar to that obtained in the ML tree (Figure 13). The phylogenetic tree of the concatenated *pheS*, *rpoA*, and *atpA* sequences proved the discriminatory power of these genes for species identification within the genus *Carnobacterium* and was roughly in agreement with the 16S rRNA gene-based analysis. All species were clearly delineated and had $<$ 98% concatenated gene sequence similarity. Pairwise sequence similarities of the concatenated *pheS*, *rpoA*, and *atpA* sequences confirmed that the closest relatives of strain LMG 26642^T were *C. funditum* LMG 14461^T (93%) and *C. mobile* LMG 9842^T (88.6%). MLSA also indicated that the closest relatives of LMG 26641 were *C. alterfunditum* LMG 14462^T (98.1%) and *C. pleistocenium* LMG 23663^T (93.8%). Lower sequence similarities ($<$ 92.5%) were found towards the other strains tested. Based on MLSA using *pheS*, *rpoA*, and *atpA*, it can be concluded that strain LMG 26641 belongs to *C. alterfunditum* and that strain LMG 26642^T can be assigned to a novel species, with *C. funditum* as its closest neighbor.

Table 5: Amplification and sequencing primers used for the MLSA of Carnobacterium *strains*

Primer name	Sequence (5'→3')	Position
pheS-24F	GAAGACTTGTTTATTGGTATGG	461
pheS-25F	CATCCAGCACGTGATATGC	557
pheS-26R	CCTAATCCAAAGGCAWATCC	1031
pheS-27R	GGATGAACVAYYCCWGCRCC	968
rpoA-24F	ATGATCGAAATTGAAAAACC	1
rpoA-25R	ACTTTAATCATTTCTGCTTC	844
rpoA-26R	ACTGAATTAATACCAGCACG	802
atpA-29F	TATRTCGGAGACGGGATTGC	97
atpA-30F	GCNCCWGGMGTWATGSAWCG	397
atpA-33R	GCTTGTTTYGWTAARTCATC	781
atpA-34R	TTCATSGCTTTAATTTGWGC	1108
atpA-35R	CCACGATTTAATTTAGCTTG	1219

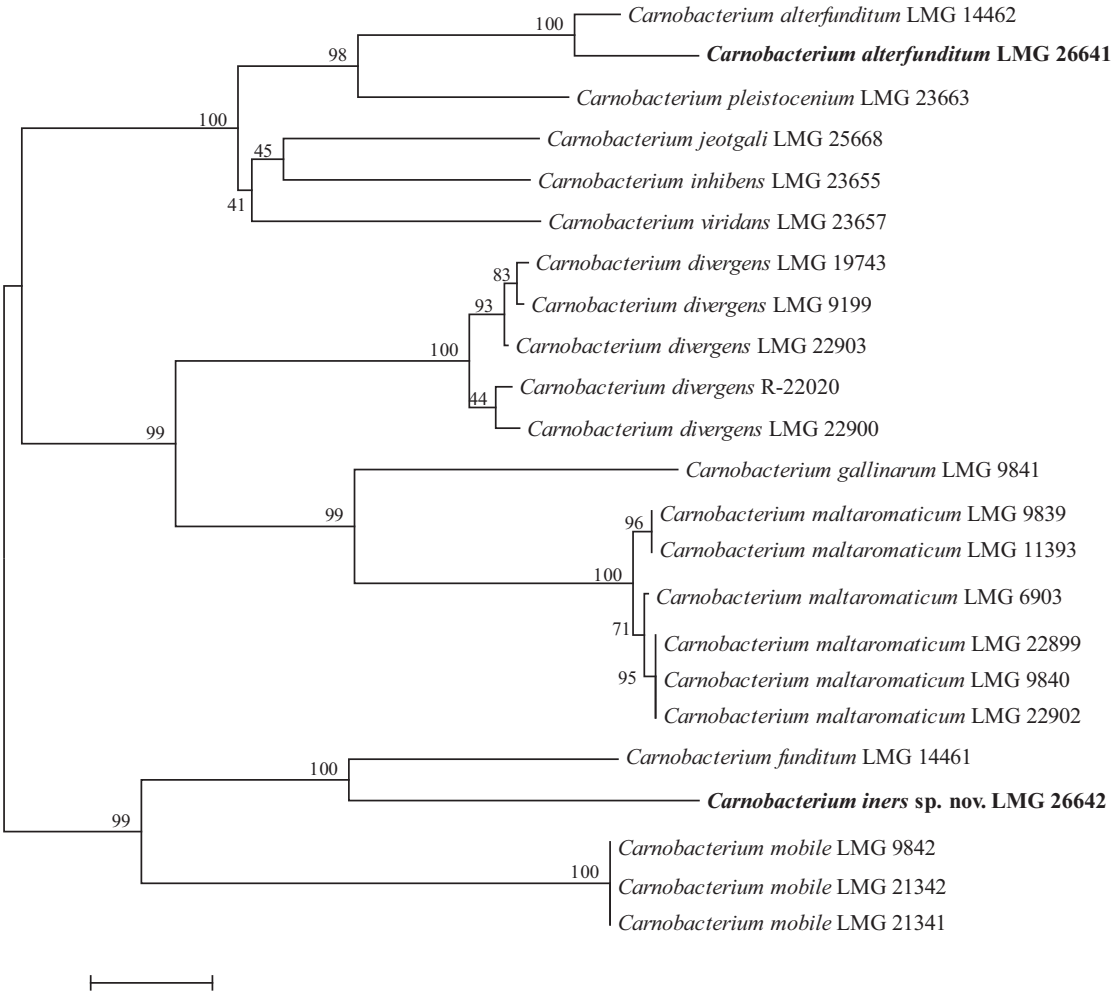


Figure 13: ML tree based on concatenated pheS, rpoA, *and* atpA *sequences showing the phylogenetic relationship of strains LMG 26642^T and LMG 26641 with the type strains of species of the genus* Carnobacterium.

Bootstrap values (> 50%) based on 1000 replications are shown at branch nodes. Accession numbers are given in Table 6. Bar, 2% sequence divergence. The substitution model used was the GTR model and the aligned sequence had a length of 957 bp.

Table 6: Members of the genus Carnobacterium *examined in this study*

Strain	Depositor*	Source	Growth conditions	Accession numbers		
				pheS	*rpoA*	*atpA*
C. alterfunditum						
LMG 14462^T	DSMZ	Anoxic water (Antarctica)	TSBY salt medium, 20°C, A	HE590697	HE590698	HE590699
LMG 26641	Own isolate	Microbial mat (Antarctica)	TSBY salt medium, 20°C, A	HE590759	HE590760	HE590761
C. divergens						
LMG 9199^T	NCIMB	Vacuum-packed refrigerated minced beef (South-Africa)	CBA, 37°C, MA	HE590700	HE590701	HE590702
LMG 22903	J. Leisner	Spoiled pork (Denmark)	CBA, 37°C, MA	HE590703	HE590704	HE590705
LMG 22900	J. Leisner	Beef (Denmark)	CBA, 37°C, MA	HE590706	HE590707	HE590708
R-22020	J. Leisner	Vacuum packed pork (ND)	CBA, 37°C, MA	HE590709	HE590710	HE590711
LMG 19743	J. Leisner	Modified-atmosphere-packaged shrimps (Denmark)	CBA, 37°C, MA	HE590712	HE590713	HE590714
C. funditum						
LMG 14461^T	DSMZ	Anoxic water (Antarctica)	TSBY salt medium, 20°C, A	HE590696	HE590715	HE590716
C. gallinarum						
LMG 9841^T	NCFB	Ice slush around chicken carcasses (ND)	MRS agar, 30°C, A	HE590717	HE590718	HE590719
C. inhibens						
LMG 23655^T	DSMZ	Gastrointestinal tract of an Atlantic salmon (Sweden)	TSA, 30°C, A	HE590720	HE590721	HE590722
C. jeotgali						
LMG 25668^T	JCM	Jeotgal (Korea)	TSA, 30°C, A	HE590723	HE590724	HE590725
C. maltaromaticum						
LMG 22899	J. Leisner	Modified-atmosphere packaged thawed cod (Denmark)	TSA, 30°C, A	HE590726	HE590727	HE590728
LMG 11393	DSMZ	Vacuum-packed beef (ND)	TSA, 30°C, A	HE590729	HE590730	HE590731
LMG 6903^T	NCIMB	Raw milk (ND)	TSA, 30°C, A	HE590732	HE590733	HE590734
LMG 9839	NCFB	Kidney of a diseased adult cutthroat trout (USA)	TSA, 30°C, A	HE590735	HE590736	HE590737
LMG 22902	J. Leisner	Modified atmosphere packaged beef (Denmark)	TSA, 30°C, A	HE590738	HE590739	HE590740
LMG 9840	NCFB	Refrigerated vacuum-packaged meat (ND)	TSA, 30°C, A	HE590741	HE590742	HE590743
C. mobile						
LMG 9842^T	NCFB	Irradiated chicken meat (ND)	TSA, 30°C, A	HE590744	HE590745	HE590746
LMG 21341	J. Leisner	Modified-atmosphere-packaged shrimps (Denmark)	TSA, 30°C, A	HE590747	HE590748	HE590749
LMG 21342	J. Leisner	Modified-atmosphere-packaged shrimps (Denmark)	TSA, 30°C, A	HE590750	HE590751	HE590752
C. pleistocenium						
LMG 23663^T	DSMZ	Ice core sample of a pleistocene thermokarst pond (Alaska)	CBA, 20°C, A	HE590753	HE590754	HE590755
C. viridans						
LMG 23657^T	DSMZ	Vacuum-packed ham (Canada)	TSA, 30°C, A	HE590756	HE590757	HE590758
C. iners sp. nov.						
LMG 26642^T	Own isolate	Microbial mat (Antarctica)	TSBY salt medium, 4°C, A	HE578182	HE578183	HE578184

CBA, Columbia blood agar (Oxoid); TSA, tryptone soya agar (Oxoid); A, aerobic; mA, microaerobic, ND, no data available.

* RIKEN BioResource Center, Japan; J. Leisner, Department of Veterinary Disease Biology, Faculty of Life Sciences, University of Copenhagen, Denmark.

To confirm the identification obtained by MLSA, DDH between strain LMG 26642^T and its nearest phylogenetic neighbor, *C. funditum* LMG 14461^T, was performed. Genomic DNA was extracted using the guanidinium thiocyanate method described by Pitcher *et al.* (1989). DDH was performed using the microplate method, with photobiotin for labeling the DNA (Ezaki *et al.*, 1989), as modified by Goris *et al.* (1998). DNA-DNA relatedness between strain LMG 26642^T and *C. funditum* LMG 14461^T was 18%, which confirmed that strain LMG 26642^T represents a novel species.

The G+C content was determined according to the enzymatic DNA degradation method described by (1989), using a Waters Breeze HPLC system and XBridge Shield RP18 column. The solvent used was 0.02 M $NH_4H_2PO_4$ (pH 4.0)/1.5% (v/v) acetonitrile. Non-methylated lambda phage DNA (Sigma) was used as the calibration reference and *Escherichia coli* LMG 2093 DNA was included as the control. The G+C content of strain LMG 26642^T was 34 mol%, which is within the range for the genus *Carnobacterium* (32-44 mol%). Strain LMG 26641 had a G+C content of 36 mol%, which corresponded with the G+C content of *C. alterfunditum* (32-36 mol%).

Preparation of the peptidoglycan and analysis of its structure were performed according to Schumann (2011). The diagnostic *meso*-diaminopimelic acid was found in addition to alanine and glutamic acid in the total hydrolysate (4 M HCl, 100°C, 15 h). The partial hydrolysate (4 M HCl, 100°C, 15 min) contained the peptides L-Ala-D-Glu and diaminopimelic acid-D-Ala. From these data, it was concluded that strain LMG 26642^T contains the peptidoglycan type A1γ.

Physiological and biochemical tests were performed using strain LMG 26642^T. Cell and colony morphology and the presence of catalase and oxidase was investigated after growth on TSBY salt agar for 10 days, incubated aerobically at 4°C. Standard biochemical tests and growth experiments were determined in TSBY salt broth and performed in triplicate, unless stated otherwise. Growth was tested at 4, 15, 20, and 37°C and in the presence of 2-10% (w/v) NaCl (at intervals of 2%). Production of gas from glucose was tested using inverted Durham tubes and peptone yeast extract as the basal medium (Holdeman & Moore, 1977). The production of D- and L-lactate from glucose was determined enzymatically (R-Biopharm). The API 50 CHL *Lactobacillus* identification system (bioMérieux) and GEN III Omilog ID system (Biolog) were used to determine the carbohydrate fermentation profile of the isolate and a positive control (*C. funditum* LMG 14461^T). For both identification systems, strain LMG 26642^T scored negative for all carbohydrates tested. The

same result was obtained when peptone yeast extract medium was used as the basal medium (Holdeman & Moore, 1977). The morphological and physiological characteristics of the isolate are given in the species description and an overview of the physiological differences between the isolate and its closest relatives is presented in Table 7.

Table 7: Differential characteristics of strain LMG 26642^T and members of the genus Carnobacterium *that assimilate a limited range of carbohydrates*

Characteristic	1	2	3
Acid from:			
Galactose	-	w	-
Mannitol	-	+	+
Salicin	-	+	ND
Glycerol	-	w	-
D-Cellobiose	-	+	-
D-Fructose	-	+	+
Maltose	-	+	-
D- Mannose	-	+	-
D-Trehalose	-	+	-
D-Ribose	-	+	-
Sucrose	-	+	-
Hydrolysis from aesculin	+	w	+
Lactic acid configuration	L	L(+)	ND
Growth in:			
2% NaCl	+	ND*	+
4% NaCl	-	+	+
DNA G+C content (mol%)	34	32-34	43.9

Strains: 1, *C. iners* sp. nov. LMG 26642^T; 2, *C. funditum* LMG 14461^T (Franzmann *et al.*, 1991); 3, *C. jeotgali* LMG 25668^T (Kim *et al.*, 2009). +, Positive; W, weakly positive; -, negative. * Sodium is required for growth with optimal growth at 1.7%

4.3.3 Conclusion

The colony and cell morphology of strain LMG 26642^T, its motility, catalase activity, peptidoglycan type, and G+C content were in agreement with the description of the genus *Carnobacterium*. Additionally, strain LMG 26642^T was closely related to *C. funditum* LMG 14461^T (99.2% 16S rRNA gene sequence similarity) but it could be distinguished from this reference strain and other members of the genus *Carnobacterium* on the basis of MLSA and biochemical characteristics. Based on these results, we propose to classify strain LMG 26642^T in a novel species, *Carnobacterium iners* sp. nov.

Description of Carnobacterium iners sp. nov.

Carnobacterium iners sp. nov. (in'ers. L. neut. adj. iners inactive, lazy). Cells are psychrophilic, Gram-stain-positive, catalase-negative, facultatively anaerobic, and motile rods, approximately 1.5 μm wide and 3-6 μm long, occurring singly or in pairs or short chains. Colonies grown for 10 days on TSBY salt agar at 4°C are approximately 0.8 mm in diameter, white, opaque, smooth and circular with undulate margins and an umbonate elevation. No gas is produced from glucose. Produces D-and L-isomers of lactic acid in a ratio of 1:9. Optimal growth is observed at 4°C; grows at 15 and 20°C. Grows with 1-2% NaCl, but not with 4-10% NaCl. Does not produce acid from glucose, fructose, mannose, N-acetylglucosamine, aesculin, cellobiose, maltose, trehalose, gentiobiose, glycerol, erythritol, D- or L-arabinose, ribose, D- or L-xylose, adonitol, methyl β-D-xylopyranoside, galactose, sorbose, rhamnose, dulcitol, inositol, mannitol, sorbitol, methyl α-D-mannopyronoside, methyl α-D-glucopyranoside, amygdalin, arbutin, salicin, lactose, melibiose, sucrose, inulin, melezitose, raffinose, starch, glycogen, xylitol, turanose, D-lyxose, D-tagatose, D- or L-fucose, D- or L-arabitol, potassium gluconate, potassium 2-ketogluonate, or potassium 5-ketogluconate. The cell wall contains *meso*-diami-nopimelic acid.

The type strain, LMG 26642^T (=CCUG 62000^T), was isolated from a cyanobacterial mat growing in the littoral zone of a continental Antarctic lake (Forlidas Pond, Pensacola mountains) in December 2003. The DNA G+C content of the type strain is 34 mol%.

4.3.4 Acknowledgements

We acknowledge financial support from the Research Foundation-Flanders and the Belgian Science Policy Office (BelSPO AMBIO project). We thank Dominic Hodgson and Peter Convey (British Antarctic Survey) who provided the environmental sample from which the isolate was obtained.

4.3.5 Supplementary Material

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene, *pheS*, *rpoA*, and *atpA* sequences of strain LMG 26642^T are HE583595 and HE578182-HE578184, respectively.

5

Microbial Diversity Analyses in the HT Sequencing Era

This Chapter comprises the application of HT sequencing technologies to unravel the microbial diversity in an acidic beer ecosystem, along with a whole-genome sequence analysis of a LAB strain, isolated from this ecosystem. In Section 5.1, a HT sequencing approach, comprising targeted sequencing of the bacterial partial 16S rRNA gene and the fungal ITS region was applied to study the ecosystem of mature Belgian red-brown acidic ales. Belgian red-brown acidic ales are sour and alcoholic fermented beers, which are produced by mixed-culture fermentation. *Saccharomyces* dominates the ethanolic fermentation phase, after which lactobacilli produce lactic acid. *Dekkera* species, pediococci, and AAB prevail during the maturation phase, which takes place in oak barrels for about two years, after which the mature beer is blended with young, non-aged beer. The goal was to compare the microbial community diversity and metabolite composition of three Belgian red-brown acidic ales. Following, whole-genome sequencing of *Pediococcus damnosus* LMG 28219, which was the dominant LAB member of the Belgian red-brown acidic ale ecosystem, and comparative genome analysis were used in Section 5.2 to investigate this strains' mechanisms to reside in the beer niche.

5.1 Microbial Diversity and Metabolite Composition of Belgian Red-Brown Acidic Ales

I. Snauwaert, S. P. Roels, F. Van Nieuwerburg, A. Van Landschoot, L. De Vuyst, P. Vandamme

Submitted to: Applied and Environmental Microbiology, 2014

Author contributions: IS carried out all experiments and performed bioinformatics analyses. SPR performed the statistical analysis. FVN participated in the 454 pyrosequencing. AVL was involved in the collection of the samples. IS, LDV, and PV wrote the manuscript. All authors read and approved the final manuscript.

Belgian red-brown acidic ales, encompassing both red and red-brown acidic ales, are sour and alcoholic fermented beers, which are produced by mixed-culture fermentation. These beers are aged in oak barrels for about two years, after which mature beer is blended with young, non-aged beer to obtain the end-products. The microbial ecology of a Belgian red acidic ale has previously been characterized, setting the foundations of current knowledge on mixed-culture fermentations. The objective was to evaluate the microbial community diversity of Belgian red-brown acidic ales at the end of the maturation phase of three breweries and of three fermentation batches per brewery. The microbial diversity is compared with the metabolite composition at the end of the maturation phase. In this work, brew samples were subjected to 454 pyrosequencing of the 16S rRNA gene (bacteria) and the ITS region (yeasts) and a broad range of metabolites was quantified. The most important microbial species present in the Belgian red acidic ales were *Pediococcus damnosus*, *Dekkera bruxellensis*, and *Acetobacter pasteurianus*. In addition, culture-independent analysis revealed operational taxonomic units (OTUs) that were assigned to an unclassified fungal community member, *Candida*, and *Lactobacillus*. The main metabolites were L-lactic acid, D-lactic acid, and ethanol, whereas acetic acid was produced in lower quantities. The most prevailing aroma compounds were ethyl acetate, isoamyl acetate, ethyl hexanoate, and ethyl octanoate, which might be of impact on the aroma of the end-products.

5.1.1 Introduction

Beer production is generally a monoculture fermentation process, with the yeast *Saccharomyces cerevisiae* as the most important player. However, the fermentation and/or maturation of some Belgian acidic beers is determined by autochthonous nonstarter microorganisms or uncharacterized mixed starter cultures (Verachtert & Iserentant, 1995; Verachtert & Derdelinckx, 2005). These beers represent culturally important alcoholic beverages, for which the microbial activities play critical roles not only in their production but also regarding quality characteristics. Indeed, the end-products of these fermentations have a unique taste, for which they are gaining popularity worldwide.

Well-known Belgian acidic beers that are the result of the activities of an autochthonous microbiota are the lambic beers, which are characterized by a spontaneous fermentation process involving enterobacteria (the first phase), *S. cerevisiae*, and *Saccharomyces pastorianus* (the main fermentation phase), and *Dekkera bruxellensis* and *Pediococcus damnosus* (the maturation phase) (Verachtert & Iserentant, 1995; Verachtert & Derdelinckx, 2005; Spitaels *et al.*, 2014b). Other examples are the ACAs, for which production methods similar to those of lambic beers are used (Bokulich *et al.*, 2012a). The maturation phase of the production of these beers is dominated by the yeast *D. bruxellensis*, which produces a range of flavor compounds, including octanoic acid and decanoic acid and their respective ethyl esters (Bokulich & Bamforth, 2013). LAB are the second most prominent group of microorganisms in these beers, whose activities result in increased lactic acid concentrations.

Much less studied however are the Belgian red-brown acidic ales, which are produced by a mixed-culture fermentation (Martens *et al.*, 1997). Nowadays, these beers are produced by only a few breweries in West- and East- Flanders (Belgium) that produce these beers for years and by some American craft breweries that aim to produce a comparable style. Two types of Belgian red-brown acidic ales exist, namely the red acidic ales of South-West-Flanders and the red-brown acidic ales that are produced in South-West- and South-East-Flanders.

Martens and colleagues monitored the microbial communities during the industrial production of a Belgian red acidic ale and described a tandem fermentation process (1997). *Saccharomyces* dominates the ethanolic fermentation phase, after which lactobacilli start the production of lactic acid. *Dekkera* species, pediococci, and AAB prevail during the maturation phase, which takes place in oak barrels.

After two years of aging, the pH drops to 3.2-3.5 and lactic acid and acetic acid concentrations increase to 6000 mg/L and 1600 mg/L, respectively, are reached (Martens *et al.*, 1997). Finally, the end-product is obtained by blending the sour mature beer and young beer that did not undergo aging in oak vessels.

The current study attempted to update and extend the existing data on Belgian red-brown acidic ales by reanalyzing their microbial diversity and metabolite composition at the end of the maturation phase. In this study the microbial diversity and metabolite composition of Belgian red-brown acidic ales are compared.

5.1.2 Materials and Methods

The Experimental Material

Three subsequent brews (samples referred to as A, B, and C) of three Belgian red-brown acidic ale breweries (referred to as no. 1, 2, and 3) were sampled at the end of their maturation phase. Red acidic ales and red-brown acidic ales were produced by brewery 1 and breweries 2 and 3, respectively. All microorganisms present were collected through microcentrifugation (8000 rpm, 20 min, 4°C) and the cell pellets were washed with resuspension buffer (0.15 M NaCl, 0.1 M ETDA, pH 8.0). The cell pellets and supernatants were stored separately at -20°C until further analysis. In addition, samples of the bottled beers produced from these brews were stored in triplicate at -20°C.

Microbial Diversity Analysis of the Brew Samples

Total DNA Extraction

Total bacterial and fungal DNA was extracted using the procedure described by Gevers and colleagues (2001), with several modifications, as follows. (i) The thawed pellet was washed in 1 ml TES buffer (6.7% sucrose, 50 mM Tris-HCl, pH 8.0, 1 mM EDTA) and resuspended in 275 μL STET buffer (8.0% sucrose, 5.0% Triton X-100, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA). (ii) Ninety μL lysozyme-mutanolysin-proteinase K solution (TES buffer containing 1667 U/mL mutanolysin, 33 mg/mL lysozyme, and 2.78 mg/mL proteinase K) was added and the suspension was incubated at 37°C for 1 h. (iii) Prior to extraction with a phenol/chloroform/isoamyl

alcohol (49.5:49.5:1.0) solution (Sigma-Aldrich, St. Louis, MO, USA), proteins were precipitated by adding 250 μL ammonium acetate. (iv) Five μL RNase (2 mg/L) was added to the DNA solution, which was incubated at 37°C for 60 min and subsequently stored at -20°C. The integrity, concentration, and purity of the DNA isolated was evaluated using 1% (wt/vol) agarose gels, stained in ethidium bromide, and by spectrophotometric measurements at 234, 260, and 280 nm. Per brew sample, DNA extracts were prepared in triplicate and pooled. Pooled DNA solutions were diluted to a concentration of 50 ng/μL prior to amplification.

Multiplex 454 Pyrosequencing of a Bacterial and Fungal Marker

Amplification was performed using a two-step PCR procedure that increases the reproducibility and consistently recovers a higher genetic diversity in pyrosequencing libraries (Berry *et al.*, 2011). To analyze the bacterial communities, a 16S rRNA gene fragment encompassing the variable regions V1 and V3 was amplified using the forward primer pA 5'-AGAGTTTGATCCTGGCTCAG-3' (Sigma-Aldrich) and the reverse primer 518R 5'-ATTACCGCGGCTGCTGG-3' (Sigma-Aldrich). The thermal program used for the amplification of the V1-3 region of the 16S rRNA gene started with a denaturation at 95°C for 5 min, followed by 20 cycles of amplification (45 s at 95°C, 45 s at 55°C, and 30 s at 72°C), and a final elongation at 72°C for 8 min. The fungal ITS region was amplified using the forward primer ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3'; Sigma-Aldrich) (Gardes & Bruns, 1993) and the reverse primer ITS4 (5'-CCTCCGCTTATTGATATGC-3'; Sigma-Aldrich) (White *et al.*, 1990). The following thermal program was applied for fungal DNA amplification: denaturation at 95°C for 5 min, 3 cycles of pre-amplification (1 min at 95°C, 2 min 15 s at 55°C, and 1 min 15 s at 72°C) followed by 30 amplification cycles (1 min at 95°C, 1 min 15 s at 55°C, and 1 min 15 s at 72°C), and elongation at 72°C for 7 min. All PCR assays were conducted in triplicate, using the following reagents: 1.25 U/reaction FastStart High Fidelity Enzyme blend (Roche Diagnostics, Basel, Switzerland), GeneAmp® dNTP's (200 μM of each; Life Technologies, Carlsbad, CA, USA), 0.4 μM of the forward and reverse primers, and FastStart High Fidelity Reaction buffer containing 1.8 mM $MgCl_2$ (Roche Diagnostics). The Veriti® 96-Well Fast Thermal Cycler (Life Technologies) was used for all amplification reactions. The PCR products were evaluated using 1% (wt/vol) agarose gels, stained in ethidium bromide, and they were pooled per brew sample afterwards. Subsequently, sequencing constructs (primers of the second PCR step) of the GS FLX Titanium chemistry were added to the amplicons, using the thermal program and PCR mixture applied before, with the following modifications. (i) The number of cycles for the pre-amplification and amplification runs was reduced to one

and five, respectively. (ii) The sequencing constructs were obtained from Integrated DNA technologies (IDT, Leuven, Belgium), as recommended by Roche Diagnostics, and their concentrations in the PCR mixture were reduced to 0.1 μM. The constructs are compatible with Lib-L chemistry and consisted of a Titanium adaptor, an eight base pair (bp) multiplex identifier (MID), and a template-specific PCR primer. The PCR products were purified using a MinElute PCR Purification kit (Qiagen, Venlo, The Netherlands) and checked using the 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with the DNA 7500 kit (Agilent Technologies). A fluorescent stain base kit (Qubit® dsDNA Broad range assay kit; Life Technologies) was used for DNA quantification. Subsequently, the PCR products of different samples were pooled in equimolar concentrations and sequenced unidirectionally on a GS FLX instrument (half run), using Titanium chemistry (Roche Diagnostics), at Beckman Coulter Genomics SA (Grenoble, France). The 16S rRNA gene amplicons were sequenced in the reverse direction, whereas the fungal ITS amplicons were sequenced in the forward direction. A mock community containing both bacterial as well as fungal DNA and a generous donor sample (*i.e.*, a sample that was sequenced before) were included in the analysis to enable error analysis.

Processing of the Sequence Data

The raw .sff 454 pyrosequencing file was processed using a 16S rRNA gene pipeline constructed by Schloss and coworkers (`http://www.mothur.org/wiki/454_SOP`) (2011a), with several modifications, as follows. (i) No nucleotide mismatches to the sample barcode and adjacent PCR primer were allowed for sample multiplex barcode deconvolution and primer trimming. (ii) Sequences with an ambiguous base call or a homopolymer stretch longer than six nucleotides were removed from subsequent analysis. (iii) The Mothur implementation of the SeqNoise algorithm was used to further reduce the overall error rate (Quince *et al.*, 2011). Additional modifications were needed for the analysis of the fungal ITS region. (i) The full ITS1 of the nuclear ITS region was extracted using the V-Xtractor package (Hartmann *et al.*, 2010) and fungal HMMs developed by Nilsson and coworkers (2010). (ii) Sequences were aligned to each other using the needleman alignment method (a gap was only penalized once, terminal gaps were penalized, and all distances were calculated), and uncorrected pairwise distances between sequences were calculated. (iii) An in-house developed training set, containing ITS sequences retrieved from the YeastIP database (Weiss *et al.*, 2013), was used for taxonomic assignments. All identifications were reviewed by means of the Mycobank identification tool (Crous *et al.*, 2004).

The bacterial and fungal datasets were normalized (normalized value = relative abundance * norm; norm = the number of sequences in the smallest group) and binned into OTUs with a cut-off of 97%, and rarefaction analyses were performed. The community diversity was measured using the inverse Simpson diversity index (1/D). Furthermore, differences in community structure and membership were not only explored using the Bray-Curtis and Jaccard index, but also using weighted and unweighted UniFrac (Lozupone *et al.*, 2007). These differences were quantified by means of analysis of similarity (ANOSIM) and visualized using principal coordinate analysis (PCoA). The R Studio software package (version 0.97.551) was used for heatmap construction and PCoA.

Isolation of Dominant Community Members from Brew Samples

A mature brew sample from brewery 1 was serially diluted in 0.86% (wt/vol) saline and 50 μL of each dilution was plated on different agar media in triplicate, followed by incubation at 28°C in an aerobic atmosphere for 4-6 days. The dominant LAB and AAB community members were isolated using MRS (Oxoid) and mDMS (Papalexandratou *et al.*, 2013) agars, respectively, supplemented with 5 ppm amphotericin B (Sigma-Aldrich) and 200 ppm cycloheximide (Sigma-Aldrich). DYPA [2% glucose, 0.5% yeast extract, 1% peptone, 100 ppm chloramphenicol (Sigma-Aldrich), and 1.5% agar (wt/vol)] was used as a general yeast agar isolation medium or was supplemented with 50 ppm cycloheximide (DYPAX) to favor the slow-growing *Dekkera* yeast (Abbott *et al.*, 2005; Suárez *et al.*, 2007). Isolates were subcultured twice and subjected to MALDI-TOF MS for dereplication, as described previously (Ghyselinck *et al.*, 2011). As sequence analysis of protein-encoding genes has proven to be superior for accurate species level identification of a variety of LAB and AAB (Cleenwerck *et al.*, 2010; De Bruyne *et al.*, 2007; Naser *et al.*, 2005a; Snauwaert *et al.*, 2013a), representative bacterial isolates were identified through sequence analysis of the *pheS* gene or the molecular chaperone (*groEL*) gene, for *Pediococcus* and *Acetobacter*, respectively, in addition to sequencing of the 16S rRNA gene. Representative fungal isolates were identified by sequencing of the ITS region and D1/D2 region of the 26S rRNA gene (Kurtzman & Robnett, 1998).

Metabolite Target Analysis of the Brew Samples and Bottled Beers

Determination of Carbohydrate Concentrations

High-performance anion exchange chromatography (HPAEC) with pulsed amperometric detection was used to determine the concentrations of glucose, fructose, sucrose, and mannitol. An ICS 3000 chromatograph (Dionex, Sunnyvale, CA, USA) with a CarboPacTM PA10 column (Dionex) was used for this purpose. The mobile phase, at a flow rate of 1.0 ml/min, consisted of ultrapure water (eluent A), 167 mM NaOH (eluent B), and 500 mM NaOH (eluent C), with the following gradient: 0.0 min, 87% A and 13% B; 20.0 min, 87% A and 13% B; 25 min, 57% A, 13% B, and 30% C; 26 min, 100% C; 30 min, 100% C; 31 min, 87% A and 13% B; and 35 min, 87% A and 13% B (Janssens *et al.*, 2012). To remove proteins, 150 μL of Carrez A reagent [3.6% (wt/vol) $K_4Fe(CN)_6.3H_2O$] and Carrez B reagent [7.2% (wt/vol) $ZnSO_4.7H_2O$] were added to 300 μL of the tenfold diluted samples. HPLC with an evaporative light-scattering detector (ELSD) was used for the determination of the concentrations of maltose, maltotriose, maltotetraose, maltopentaose, and maltohexaose. To this end, a Waters chromatograph (Waters Corporation, Milford, MA, USA) was used, equipped with a 2424 ELS Detector (gain, 100; drift tube, 65°C; gas pressure, 40 psi; nebulizer cooling), a 600S controller, a 717Plus autosampler, and a Grace Prevail Carbohydrate ES column (250 x 4.6 mm, 5 μm, 35°C; Grace Discovery Sciences, Deerfield, IL, USA). The mobile phase, at a flow rate of 1.0 ml/min, consisted of ultrapure water (eluent A) and acetonitrile (eluent B), with the following gradient: 0.0 min, 25% A and 75% B; 50.0 min, 40% A and 60% B; 51 min, 25% A and 75% B; and 60 min, 25% A and 75% B. Proteins were removed by adding 500 μL acetonitrile (Sigma-Aldrich) to 500 μL sample. All samples were microcentrifuged (14000 rpm for 8 min), filtered (0.2 μm filters, Sartorius AG, Göttingen, Germany), and transferred to an appropriate vessel prior to injection. Calibration was performed using external standards and samples were analyzed in triplicate.

Determination of the Concentrations of D- and L-Lactic Acid

Concentrations of D- and L-lactic acid were determined using HPLC-UV. To this end, a Waters chromatograph (Waters Corporation) was used, equipped with a 486 UV detector, a 600S controller, a 717Plus autosampler, and a Shodex-Column Orpak CRX-853 (50 x 8 mm; Showa Denko KK, Tokyo, Japan). The mobile phase, at a flow rate of 1.0 ml/min, consisted of a 1 mM $CuSO_4$ solution in ultrapure water with 20% acetonitrile (Sigma-Aldrich). D- and L-lactic acid were eluted isocratically. Detec-

tion was performed by measuring the absorption at 253 nm. Sample preparation, calibration, and number of technical replicates were as described above.

Determination of Volatiles

Static-headspace (SH)-GC-MS was used for the identification and quantification of flavor compounds. Samples were prepared in headspace vials containing 5 ml sample as well as a known amount of 1-butanol (Sigma-Aldrich) as internal standard. The same procedures and instrument control parameters were applied as reported by Ravyts and colleagues (2009), except that sample equilibration was done by agitation at 60°C for 30 min prior to injection into the column. Using this method, isobutyl alcohol, isoamyl acetate, isoamyl alcohol, ethyl hexanoate, ethyl octanoate, ethyl decanoate, 2-phenylethyl acetate, and 2-phenyl ethanol could be successfully identified. Ethyl acetate and 1-propanol could be analyzed successfully when splitting the injection port. The amount of volatiles present in the samples was calculated using external standards and all samples were analyzed in triplicate.

Determination of the Concentrations of Ethanol and Short-Chain Fatty Acids

The concentrations of ethanol and short-chain fatty acids (SCFAs) in the samples were measured by gas chromatography, using a Focus gas chromatograph (Interscience, Breda, The Netherlands) equipped with a Stabilwax-DA column (Restek, Bellefonte, PA, USA), a flame ionization detector, and an AS 3000 autosampler. Hydrogen gas was used as carrier gas with a constant flow rate of 1.0 ml/min; nitrogen gas was used as make-up gas. The injector and detector temperatures were set to 240°C and 250°C, respectively, and the following temperature program was used: 0.0 min at 40°C, 10.0 min at 140°C, 12.0 min at 230°C, and 22.0 min at 230°C. Nine hundred μL of a mixture [acetonitrile (Sigma-Aldrich), 1% (vol/vol) formic acid (Merck, Darmstadt, Germany), and 2.842 g/L 2-butanol (Merck)] was added to a 300 μL non-diluted or ten times diluted sample for the measurement of SCFAs and ethanol, respectively. Prior to injection, the sample was microcentrifuged (14000 rpm for 8 min) and filtered (0.2 μm, Sartorius AG). A volume of 1.0 μL was injected using a split ratio of 40:1. The quantification was done in triplicate using external standards, as described above.

Statistical Analysis

Missing data were excluded prior to statistical analysis of the metabolite data. A correlation matrix was calculated using the Pearson correlation coefficient. Principal component analysis (PCA) was performed to visualize possible relationships within the data matrix. Statistical analysis was performed using the R Studio software package (version 0.97.551).

A multivariate approach to unravel possible interactions between the metabolite concentrations and the microbial communities was performed, as described previously (Wang *et al.*, 2012), with several modifications. Outlying brew samples (2B and 2C) and non-informative variables (metabolites and microorganisms absent in more than 3 out of 7 samples) were excluded prior to analysis. Canonical correlation analysis (CCA) was used to examine possible correlations between individual external variables and overall microbial communities plus correlations between all external variables and each individual OTU. CCA was performed using the cca function in the Vegan package (R Studio software package, version 0.97.551); the envfit function was used to calculate the p-value of each correlation between each OTU and all variables. For this dataset, the cut-off for significance was set at $p < 0.15$. Variables and community members without any correlations were excluded from the multivariate approach. Afterwards, the selected variables were subjected to a multiple linear model (function lm and ANOVA). For the small dataset, the statistical significance was set at a p-value < 0.05. Finally, the direction and size of the correlations between variables and OTUs were derived from a linear regression model.

5.1.3 Results

Analysis of the 454 Pyrosequencing Data of the Community DNA from the Brew Samples

Sequencing of the amplified fragments yielded a total of 544,508 reads with a mean length of 507 bp, which corresponded roughly with 20,000 reads per sample. Multiplex barcode deconvolution, primer trimming, and denoising reduced the total number of bacterial and fungal sequences to 199,481 (395 unique sequences, mean read length of 245 bp) and 232,921 (475 unique sequences, mean read length of 178 bp), respectively. Mock community analysis showed that singleton OTUs represented sequencing errors and these OTUs were therefore excluded for further

analysis. After quality control, the normalized number of bacterial and fungal reads per sample ranged from 9,781 to 9,796 [mean ± standard error (SE) of 9792 ± 1.5] and 13,339 to 13,360 (13,355 ± 2.3), respectively. A total of 45 and 60 bacterial and fungal OTUs, respectively, remained, with a few OTUs representing a large fraction of all reads (see below). Rarefaction analysis of the bacterial and fungal datasets (Supplementary Material 5.1) indicated that all samples were more or less reaching the plateau phase and hence no additional sequencing was performed.

Bacterial communities predominantly consisted of the divisions *Firmicutes* (representing 81.64% of all reads, 21 OTUs) and *Proteobacteria* (18.31%, 15 OTUs). Only a minor fraction of the reads belonged to *Actinobacteria* (0.02%, 4 OTUs), *Deinococcus-Thermus* (0.01%, 1 OTU), and *Bacteroidetes* (0.02%, 4 OTUs). All *Firmicutes* reads were assigned to the order *Lactobacillales*, with the exception of 4 *Bacillales* OTUs (0.02%). The *Lactobacillales* OTUs belonged to the following genera: *Pediococcus* (74.49%, 1 OTU), *Lactobacillus* (7.66%, 8 OTUs), *Leuconostoc* (0.22%, 1 OTU), *Weissella* (0.23%, 1 OTU), and *Lactococcus* (0.01%, 2 OTUs). A major fraction of the *Proteobacteria* reads were assigned to the order *Rhodospiralles* (18.22%, 1 OTU), more specifically to a member of the family *Acetobacteraceae*. *Dekkera* was the dominant genus in the fungal dataset (59.26%, 2 OTUs), with smaller fractions of *Pichia* and *Candida* representing 7.93% (5 OTUs) and 5.72% (4 OTUs) of the fungal reads, respectively. Only one OTU was assigned to *Kregervanrija*, *Debaryomyces*, *Priceomyces*, *Hyphopichia*, and *Wickerhamomyces*, consisting of 0.24, 0.08, 0.07, 0.03, and less than 0.01% of the reads, respectively. Additionally, 26.45% (40 OTUs) of the fungal sequences remained unclassified.

Comparison of the Microbial Communities of the Brew Samples

The highest median bacterial and fungal 1/D indices were found in the brew samples of breweries 2 and 3, respectively, whereas the lowest were found in the brew samples of breweries 3 and 1, respectively. Additional distributional information can be found in the boxplots, as depicted in Figure 14. Differences in community structure were explored and visualized by means of PCoA, based on the Bray-Curtis and weighted UniFrac indices (Supplementary Material 5.2). The PCoA ordination of the bacterial dataset based on the Bray-Curtis index showed a separation of brew samples 2B and 2C from the remaining samples along the first axis. This trend was confirmed (but less pronounced) when incorporating phylogenetic distances using weighted UniFrac indices. PCoA based on the Bray-Curtis index of the fungal dataset showed that brew sample 2C and brew samples

3A and 3C were separated from the remaining samples along the first axis, whereas brew samples 2A and 2B and brew sample 3A were separated from the remaining ones along the second axis. Differences in community structure were quantified by means of ANOSIM. The community structure varied across brew samples of different breweries, exerting a significant impact on the bacterial (Bray-Curtis $R_{ANOSIM} = 0.506$, $p < 0.005$) but not on the fungal (Bray-Curtis $R_{ANOSIM} = 0.095$, $p > 0.005$) genetic diversity. Incorporating phylogenetic distances did not reveal significant differences in community structure across brew samples of different breweries. In addition, differences in community membership were explored using PCoA, based on the Jaccard and unweighted UniFrac indices (Supplementary Material 5.3). PCoA of the bacterial dataset based on the Jaccard and unweighted UniFrac indices both revealed a separation of brew samples 2A and 2B from the remaining brew samples along the first axis, whereas no separation was apparent along the second axis. PCoA of the fungal dataset based on the Jaccard and unweighted UniFrac index showed a separation of the three brew samples from brewery 2 from those from breweries 1 and 3 along the first axis. The community membership varied across brew samples of different breweries, exerting a significant impact on the bacterial (Jaccard $R_{ANOSIM} = 0.091$, $P > 0.005$) and the fungal (Jaccard $R_{ANOSIM} = 0.062$, $P > 0.005$) genetic diversity. Incorporating phylogenetic distances did not reveal significant differences in community membership across brew samples of different breweries.

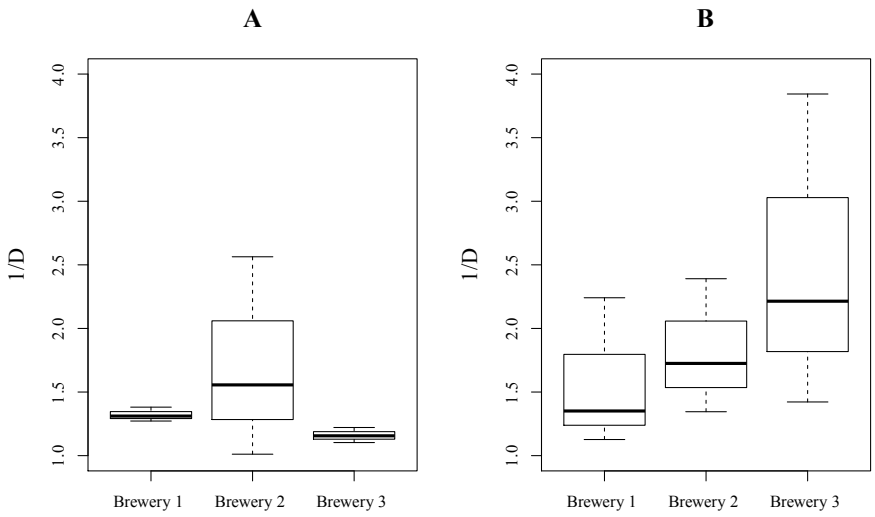


Figure 14: 1/D of the bacterial (A) and fungal (B) datasets of brew samples from breweries 1, 2, and 3

The relative abundances of the most abundant bacterial and fungal OTUs were visualized by means of a heatmap, in which the brew samples were grouped based on their Bray-Curtis distances (Figure 15 and Supplementary Material 5.4). Brew samples A, B, and C from breweries 1 and 3 comprised rather stable bacterial communities, whereas more variation was apparent in the bacterial communities of brew samples from brewery 2 (Figure 15A). The bacterial communities of all brew samples were dominated by *Pediococcus* (bacterial OTU 1), except for brew samples 2B and 2C, in which an *Acetobacteraceae* OTU (bacterial OTU 2) was predominant. Several OTUs that were assigned to the genus *Lactobacillus* represented smaller fractions. Those were OTU 5, being present in all brew samples except for brew sample 2A, and OTU 3 and OTU 4, accounting for 33.44% and 7.18% of the reads in brew sample 2B, respectively, but being less abundant in the remaining brew samples. *Weissella* (bacterial OTU 7) and *Leuconostoc* (bacterial OTU 8) were both found in brew sample 2B, whereas brew sample 3C only contained a minor fraction of OTU 7 (*Weissella*). Brew samples A, B, and C from brewery 1 comprised rather stable fungal communities, whereas more variation was apparent in the fungal communities of brew samples from breweries 2 and 3 (Figure 15B). Fungal OTU 1, which was assigned to the genus *Dekkera*, was the predominant member in the fungal communities of all brew samples, except for brew sample 2C and brew samples 3A and 3C, which were dominated by an unclassified OTU (fungal OTU 3). Two fungal OTUs were assigned to the genus *Pichia* (OTUs 4 and 5) and *Candida* (OTUs 3 and 7), respectively, whereas one OTU (OTU 10) was assigned to the genus *Kregervanrija*.

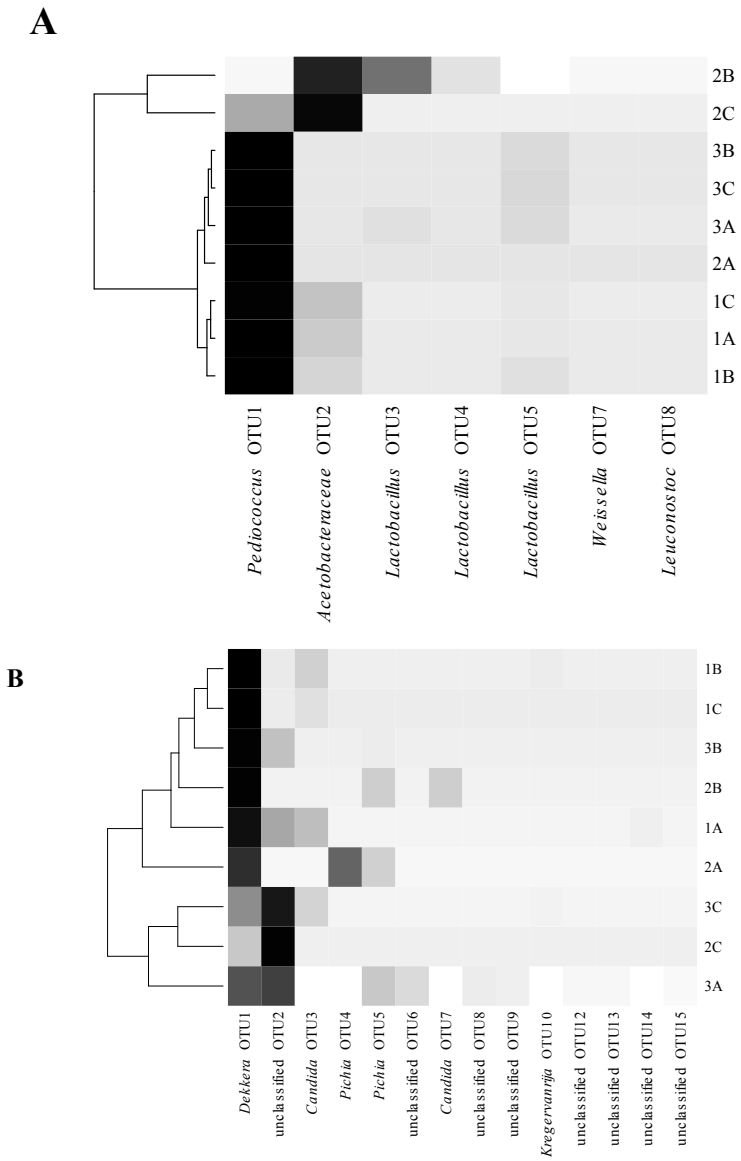


Figure 15: Relative abundances of the most abundant OTUs (> 0.01% of the reads) of the bacterial (A) and fungal (B) communities

The OTUs and sample names (breweries 1, 2, or 3; samples A, B, or C) are shown on the x- and y-axis, respectively. The color gradient ranges from black to white with decreasing relative abundances. The Bray-Curtis distances between the brew samples were calculated and an average-linkage clustering was used for the construction of the dendrograms.

Analysis of the Community Diversity Shared by the Brew Samples of the Different Breweries and Isolation of Dominant Community Members

The bacterial community members that were present in brew samples of each of the three breweries were OTU 1 (*Pediococcus*), OTU 2 (*Acetobacteraceae* unclassified), OTU 3 (*Lactobacillus*), OTU 4 (*Lactobacillus*), and OTU 5 (*Lactobacillus*). Similarly, fungal OTU 1 (*Dekkera*), OTU 2 (unclassified), OTU 4 (*Pichia*), and OTU 5 (*Pichia*) were found in all three breweries. To confirm these results and to obtain a more accurate identification of this microbiota, a targeted isolation of each of the dominant community members was conducted on a brew sample from brewery 1. Dereplication analysis of 78, 39, and 85 isolates from aerobically incubated MRS, mDMS, and DYPAX agars, respectively, yielded for these media a single MALDI-TOF MS profile (data not shown). Firstly, Sanger sequencing of the 16S rRNA and *pheS* genes assigned the taxon isolated from MRS agar to *Pediococcus damnosus*; the 251 nucleotide long stretch of the 16S rRNA gene sequence, which corresponded with the 454 pyrosequence fragment randomly picked up from bacterial OTU 1 (*Pediococcus*), shared 99.37% of its sequence. Secondly, the taxon isolated from mDMS agar was identified as *Acetobacter pasteurianus* based on 16S rRNA and *rpoB* gene sequencing; the 226 nucleotide long stretch of 16S rRNA gene sequence, which corresponded with the 454 pyrosequence fragment randomly picked up from bacterial OTU 2 (*Acetobacteraceae*), shared 100% of its sequence. Finally, Sanger sequencing of the ITS and D1/D2 regions assigned the taxon isolated from DYPAX to *Dekkera bruxellensis*; the 175 nucleotide long stretch of the ITS sequence, which corresponded with the 454 pyrosequence fragment randomly picked up from fungal OTU 1 (*Dekkera*), shared 97.74% of its sequence. No growth was obtained on DYPA medium.

Comparison of the Metabolite Data of the Brew Samples

The sucrose, fructose, and mannitol concentrations were 5.68 ± 1.79 mg/L, 154.47 ± 6.83 mg/L, and 2.27 ± 0.59 mg/L, respectively. None of the brew samples contained glucose. The concentrations of the remaining carbohydrates were 540.62 ± 48.50 mg/L maltose, 877.72 ± 72.43 mg/L maltotriose, 1760.99 ± 69.56 mg/L maltotetraose, 1342.56 ± 41.59 mg/L maltopentaose, and 1545.37 ± 42.55 mg/L maltohexaose. The average D- and L-lactic acid concentrations at the end of the maturation phase of all brew samples were 2497.35 ± 145.40 mg/L (mean ± SE) and 2175.47 ± 125.99 mg/L, respectively (details on the spread of the data are presented in Supplementary Material 5.5). The average acetic acid concentration of the whole dataset was 1376.78 ± 75.84 mg/L and only a limited number of brew

samples contained propionic acid and isobutyric acid (3.70 ± 0.86 mg/L and 9.95 ± 1.75 mg/L, respectively). The average concentrations of ethyl acetate, isobutyl alcohol, isoamyl acetate, isoamyl alcohol, and ethyl hexanoate were 151.07 ± 0.61 mg/L, 14.69 ± 0.61 mg/L, 0.88 ± 0.08 mg/L, 77.89 ± 2.92 mg/L, and 9.43 ± 1.22 mg/L, respectively. No 2-phenylethyl acetate, 2-phenyl ethanol, and only minor fractions of ethyl octanoate were found. An average ethanol concentration of 6.46 ± 0.14% (vol/vol) was measured.

To explore and visualize differences in metabolite composition among all brew samples analyzed, a PCA was performed (Figure 16A). The data points from brew samples of breweries 1 and 3 showed overlap along both axes, whereas those from brewery 2 grouped separately. Along axis 2, there was a subdivision of the data points from brew samples of brewery 2 into two clusters. In addition, the metabolite data from brew samples of breweries 1, 2, and 3 were compared by means of a heatmap (Figure 16B). The median L-lactic acid concentration was the highest in brew samples from brewery 3 and the lowest in brew samples from brewery 2, whereas the median acetic acid concentration was the highest in brewery 1 and the lowest in brewery 3. Brew samples from breweries 1 and 3 contained higher median D-lactic acid concentrations compared to those from brewery 2. The fructose concentrations were comparable in brew samples of the three breweries. The median maltose concentrations in brew samples from breweries 1 and 2 were similar, whereas the median maltose concentration in beer samples from brewery 3 was lower. Brew samples from brewery 1 contained the highest median maltotriose, maltotetraose, and maltopentaose concentrations, whereas brew samples from brewery 2 had the lowest median maltotetraose and maltopentaose concentrations. The lowest median maltotriose concentration was found in brew samples from brewery 3. No clear differences in median maltohexaose concentrations could be found. Sucrose was only found in small amounts in brew sample 3B (51.12 ± 0.62 mg/L). The median ethyl acetate concentrations were higher in brew samples from breweries 1 and 2 compared to the median ethyl acetate concentration in brewery 3. No differences in the median concentrations of isoamyl alcohol among the brew samples of the different breweries were apparent. Brew samples from brewery 3 contained less isoamyl acetate compared to the quantities found in brew samples from breweries 1 and 2. The median isobutyl alcohol and ethyl hexanoate concentrations were the highest in brew samples from brewery 2, whereas lower concentrations were found in brew samples from breweries 1 and 3. No apparent differences were found in the median concentrations of mannitol, propionic acid, and isobutyric acid.

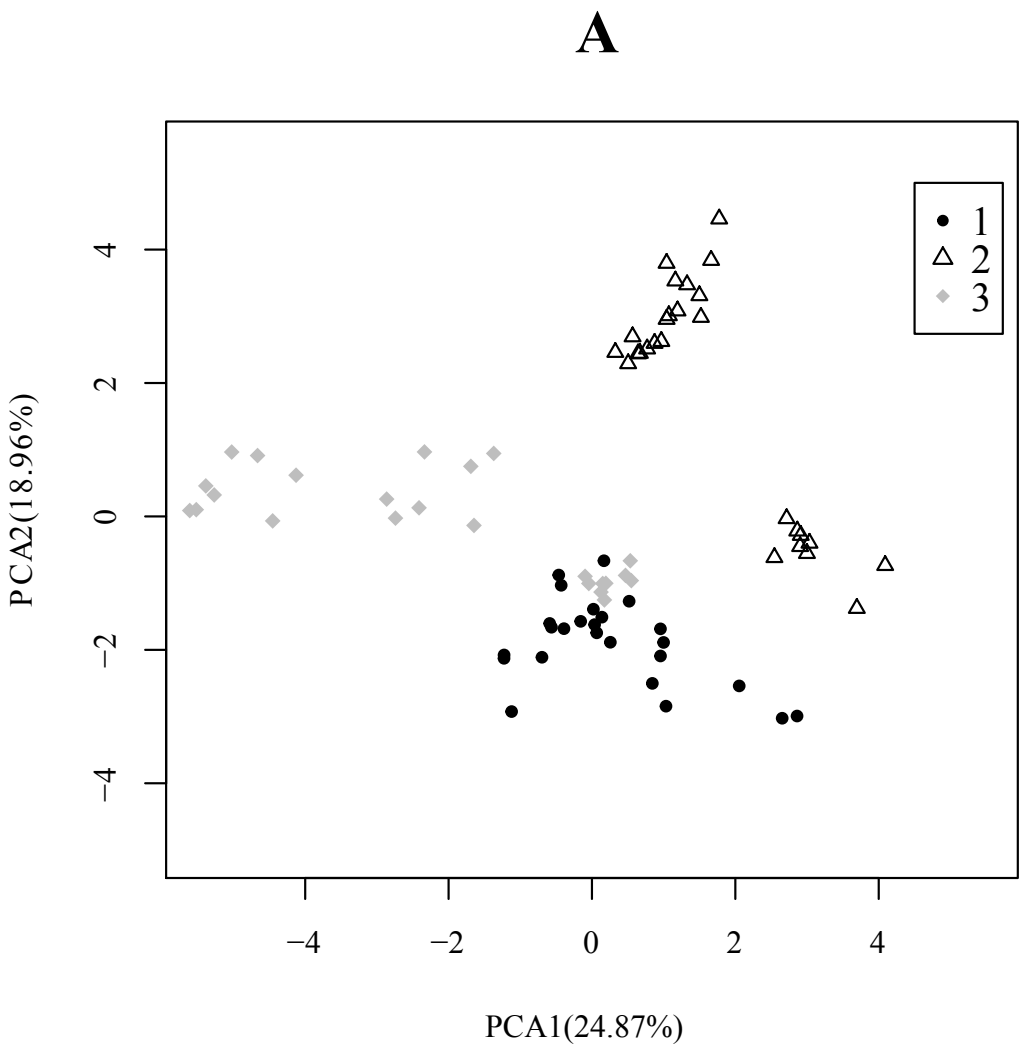


Figure 16: PCA plot (A) and heatmap (B) of the metabolite concentrations of breweries 1, 2, and 3, determined in brew samples from the end of the maturation phase

(A) In the PCA plot, the positions of the samples along the two first principal component axes are illustrated, along with the percentages of variation explained by each axis. A total of 81 data points were analyzed, including both biological and technical replicates.

Figure 16 is continued on the next page

B

		L-lactic acid	D-lactic acid	ethyl acetate	isobutyl alcohol	isoamyl acetate	isoamyl alcohol	ethyl hexanoate	ethyl octanoate	mannitol	fructose	sucrose	maltose	maltotriose	maltotetraose	maltopentaose	maltohexaose	acetic acid	propionic acid	isobutyric acid	ethanol
Brewery 1	A	2745.03	2436.39	153.88	9.88	1.17	60.95	6.04	0.06	0.00	145.64	0.00	662.53	1841.10	2513.59	1663.24	1472.07	1958.90	0.00	4.55	6.00
	B	2419.26	2447.36	161.60	12.16	1.40	68.95	4.48	0.06	0.00	172.85	0.00	688.80	1593.63	2129.07	1682.59	1275.25	2062.68	0.00	17.96	5.00
	C	2135.65	2270.74	258.21	9.88	1.14	93.16	0.00	0.06	0.00	60.08	0.00	866.34	1132.26	1356.63	1524.47	1886.33	2345.45	11.17	3.28	6.00
Brewery 2	A	2393.79	1508.18	146.37	21.74	1.29	86.18	36.40	0.08	4.81	199.27	0.00	825.27	724.63	1243.98	902.53	1015.00	1723.17	12.09	43.99	7.00
	B	711.78	876.90	25.73	20.70	0.29	69.44	18.98	0.06	0.00	160.34	0.00	736.65	988.93	1179.66	826.26	1482.70	306.31	0.00	0.00	7.00
	C	1119.75	830.33	275.34	22.99	1.70	98.31	2.93	0.06	0.00	82.97	0.00	0.00	1109.93	1284.67	1207.45	1801.76	1308.60	0.32	0.00	7.00
Brewery 3	A	5029.78	2329.56	68.96	8.76	0.00	57.56	0.00	0.00	14.69	262.08	0.00	1264.30	713.04	2916.97	1384.45	1328.79	1055.88	0.00	0.00	7.00
	B	2356.33	2166.11	82.33	12.27	0.58	59.25	0.00	0.06	0.00	170.35	0.00	0.00	0.00	1432.26	1781.04	2254.31	1047.57	0.00	13.93	7.00
	C	4489.14	5307.34	86.60	11.76	0.00	98.45	15.15	0.06	0.00	137.07	50.87	0.00	0.00	1665.40	1020.29	1244.38	493.17	2.78	0.00	4.00
		mg/l																			**v/v %**

(B) The median concentrations of all compounds are shown and colored with a gradient that ranges from black (high median concentration) to white (low median concentration). All concentrations are expressed in mg/L, except for ethanol (in vol/vol%)

Correlations Between Microbial Community Members and Metabolite Compositions

A CCA on 20 metabolites and 37 OTUs reduced the whole dataset to 6 metabolites and 9 OTUs. In total, 16 significant correlations between individual metabolites and OTUs were found through a multivariate analysis, namely 8 positive and 8 negative correlations. The fold change of the significant correlations at OTU level is depicted in Supplementary Material 5.6. The dominant *Pediococcus* (bacterial OTU 1) and *Dekkera* (fungal OTU 1) members were both negatively correlated with L-lactic acid, whereas *Pediococcus* was also negatively correlated with acetic acid but positively correlated with isobutyl alcohol. Bacterial OTU 2 (*Acetobacteraceae*) and OTU 5 (*Lactobacillus*) were negatively correlated with L-lactic acid and isoamyl acetate, respectively, whereas a positive correlation was found with isobutyl alcohol. *Pichia* OTU 4 showed a negative correlation with maltopentaose and a positive correlation with isobutyl alcohol and isobutyric acid, whereas *Pichia* OTU 5 was positively correlated with L-lactic acid. The unclassified fungal OTU 2 was negatively correlated with acetic acid, whereas OTU 29 was negatively correlated with isobutyric acid and maltopentaose. Finally, fungal OTU 24 (*Dekkera*) was positively correlated with isoamyl acetate.

Metabolite compositions of the bottled beers

The glucose, fructose, and sucrose concentrations of the bottled beers were 155.74 ± 31.17 mg/L, 2612.15 ± 227.79 mg/L, and 16.34 ± 2.77 mg/L, respectively. No mannitol could be found. The quantities of the remaining carbohydrates were 989.38 ± 74.99 mg/L maltose, 1481.94 ± 90.31 mg/L maltotriose, 3468.06 ± 96.50 mg/L maltotetraose, 1609.91 ± 64.40 mg/L maltopentaose, and 1628.55 ± 58.89 mg/L maltohexaose. The average L- and D-lactic acid concentrations of the bottled beers were 871.42 ± 48.69 mg/L (mean ± SE) and 560.57 ± 44.06 mg/L, respectively (details on the spread of the data points are presented in Supplementary Material 5.7). The average acetic acid concentration of the whole dataset was 673.66 ± 46.30 mg/L and only a limited number of samples from the bottled beers contained propionic acid (8.40 ± 1.88 mg/L). No isobutyric acid could be found. The average concentrations and SEs of ethyl acetate, isobutyl alcohol, isoamyl acetate, isoamyl alcohol, and ethyl hexanoate were 70.69 ± 4.15 mg/L, 13.64 ± 0.59 mg/L, 4.62 ± 0.28 mg/L, 78.04 ± 4.31 mg/L, and 0.69 ± 0.11 mg/L, respectively. No 2-phenylethyl acetate, 2-phenyl ethanol, and ethyl octanoate could be found. An average ethanol concentration of 4.78 ± 0.23% (vol/vol) was measured.

To explore and visualize differences in metabolite compositions, a PCA and cluster analysis were performed (Figure 17A). The data points from bottled beers from breweries 1 and 3 showed overlap along both axes, whereas those from brewery 2 showed no overlap. The metabolite compositions of the bottled beers from breweries 1, 2, and 3 were compared by means of a heatmap (Figure 17B). The median concentrations of L-lactic acid and acetic acid were the highest in bottled beers from brewery 2, whereas the median concentrations of D-lactic acid were the lowest in bottled beers from brewery 2. The highest median glucose concentration was found in bottled beers from brewery 1. High fructose concentrations were found in bottled beers from breweries 1 and 3 compared to those from brewery 2. Sucrose was only found in bottled beers from breweries 1 and 3, with average concentrations of 47.31 $\pm$ 11.08 mg/L and 47.14 $\pm$ 0.79 mg/L, respectively. No large differences were found in the median maltose concentrations across the bottled beers of the different breweries. Brewery 3 had the highest median maltotriose concentration but the lowest median maltotetraose, -pentaose, and -hexaose concentrations. The median maltotriose, -tetraose, -pentaose, and -hexaose concentrations in bottled beers from brewery 2 were slightly higher compared to those from brewery 1. The median ethyl acetate and isoamyl alcohol concentrations in bottled beers from brewery 3 were lower compared to those from breweries 1 and 2. Bottled beers from brewery 2 contained a higher median isobutyl alcohol concentration compared to those from breweries 1 and 3. No apparent differences were found in the median concentrations of isoamyl acetate, ethyl hexanoate, and propionic acid.

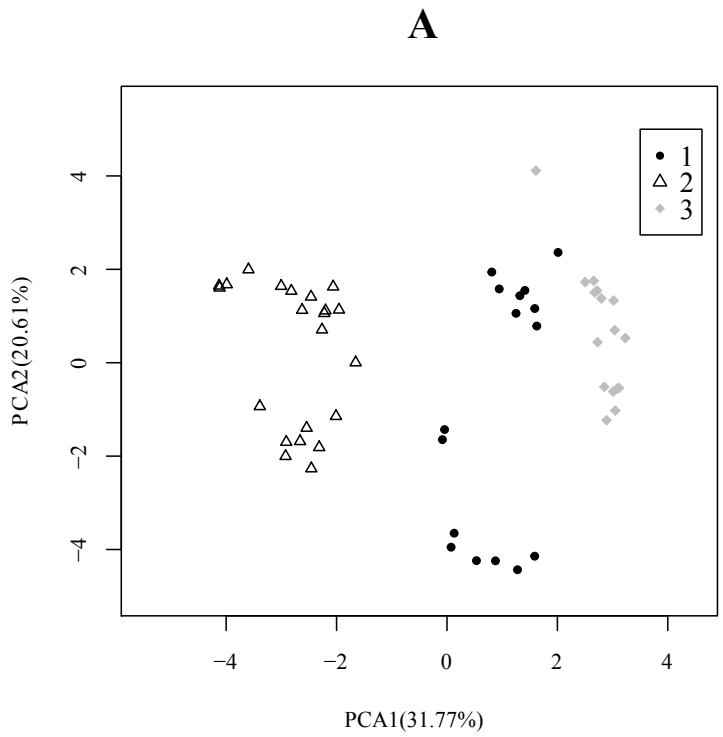


Figure 17: PCA plot (A) and heatmap (B) of the metabolite concentrations of the bottled beers from breweries 1, 2, and 3

(A) In the PCA plot, the positions of the samples along the two first principal component axes are illustrated, along with the percentages of variation explained by each axis. A total of 81 data points were analyzed, including both biological and technical replicates.

Figure 17 is continued on the next page

B

		L-lactic acid	D-lactic acid	ethyl acetate	isobuty lalcohol	isoamyl acetate	isoamyl alcohol	ethyl hexanoate	ethyl octanoate	glucose	fructose	sucrose	maltose	maltotriose	maltotetraose	maltopentaose	maltohexaose	acetic acid	propionic acid	ethanol
Brewery 1	A	550.41	434.93	44.96	10.07	3.06	64.75	2.11	0.05	784.81	5852.54	0.00	1307.60	962.09	4399.48	1739.36	1749.03	379.83	0.00	5.00
	B	368.99	398.48	81.52	10.05	1.88	88.73	0.00	0.00	601.11	5012.52	0.00	0.00	0.00	3081.70	1471.25	1643.19	553.63	51.57	5.00
	C	841.76	1047.53	108.32	10.58	3.14	149.84	0.00	0.00	0.00	3449.37	36.12	1173.84	1920.94	3692.03	1806.75	1840.93	608.01	0.00	2.00
Brewery 2	A	740.34	262.42	89.14	19.50	5.99	74.58	1.92	0.08	0.00	183.55	0.00	1356.19	1522.33	4362.03	2161.87	1887.32	1386.97	0.00	7.00
	B	1412.01	347.24	79.39	16.06	5.10	56.65	2.11	0.00	0.00	128.21	0.00	1168.93	1437.35	4098.63	2166.53	1860.15	1375.82	0.00	6.00
	C	982.38	169.96	57.00	21.29	3.78	86.92	0.00	0.00	156.18	538.42	0.00	0.00	0.00	3849.02	2223.10	2288.50	742.49	31.78	7.00
Brewery 3	A	713.33	485.43	38.21	8.05	3.93	31.27	0.00	0.00	0.00	4637.44	0.00	1096.63	2348.85	2651.69	607.27	614.16	485.67	0.00	5.00
	B	783.24	1020.20	55.18	14.11	6.34	74.74	0.00	0.00	0.00	2278.57	47.00	2037.26	2108.86	2645.38	1295.17	1565.89	295.04	0.00	2.00
	C	1165.87	1417.25	38.32	12.10	4.95	61.02	0.00	0.00	0.00	2402.66	48.58	564.03	1670.10	2087.77	937.20	1050.42	205.82	3.43	2.00
		mg/l																		**v/v %**

(B) The median concentrations of all compounds are shown and colored with a gradient that ranges from black (high median concentration) to white (low median concentration). All concentrations are expressed in mg/L, except for ethanol (in vol/vol%)

5.1.4 Discussion

The present study reanalyzed the microbial species diversity and metabolite composition of a Belgian red acidic ale previously characterized by traditional, culture-dependent techniques along with limited chemical analyses (Martens *et al.*, 1997), in comparison with two other Belgian red-brown acidic ales. Each of these beers was produced through mixed-culture fermentation and maturation in oak barrels (Martens *et al.*, 1997).

Viewed at higher taxonomic levels (*i.e.*, judging from partial 16S rRNA gene and ITS1 region sequence analysis), the microbial diversity of the brew samples of brewery 1 (*i.e.*, red acidic ale) was similar to that reported by Martens and coworkers (1997). Indeed, *Dekkera* and *Pediococcus* predominated during red acidic ale maturation, with *Dekkera* being mainly responsible for super-attenuation of the beer, *i.e.*, the depletion of oligo- and polysaccharides, and *Pediococcus* for lactic acid production. This coexistence of yeasts and LAB at the end of the maturation phase is a known phenomenon during gueuze beer fermentation (Spitaels *et al.*, 2014b; Van Oevelen *et al.*, 1977; Kumara *et al.*, 1993). Also, previously reported mixed culture experiments indicate that super-attenuation by *Dekkera* is less pronounced when pediococci are absent and suggest that growing cells of *Dekkera* can hydrolyse the EPSs produced by pediococci (Kumara & Verachtert, 1991; Van Oevelen & Verachtert, 1979). Besides these dominant members, the present study revealed the presence of an *Acetobacteraceae* member, an unclassified fungus, *Candida* and *Lactobacillus* in the brew samples of this brewery. The presence of AAB agrees with the findings of Martens and coworkers (1997), who isolated a minor fraction of these bacteria and probably reflects the presence of oxygenated niches in the vertical maturation casks. As the unclassified fungus was most abundant in one brew sample only, further research is needed to reveal its identity and function. The presence of *Candida*, which is known to be capable of growing in beer and to occur at the air-beer interface (van der Aa Kuhle & Jespersen, 1998; Timke *et al.*, 2008), contrasted with the previously reported findings for Belgian red acidic ales (Martens *et al.*, 1997), in which *Candida* species could not be recovered at the end of the maturation phase. Finally, the red acidic ale microbiota of this brewery also comprised a *Lactobacillus*. Indeed, several lactobacilli reside in beer (Bokulich & Bamforth, 2013) and were recovered during primary and secondary fermentation of the previously studied red acidic ale, but never from the mature product (Martens *et al.*, 1997).

At present, HT sequencing cannot reliably differentiate many closely related bacterial species because of the limited read length of HT sequencing technologies. Hence, HT sequencing cannot be considered fully comprehensive if detailed species level dynamics must be revealed, which is common practice in the study of food fermentations (Bokulich & Mills, 2012d). Therefore, the targeted isolation of microorganisms of interest, which were previously revealed by HT sequencing, is an appealing strategy for subsequent in-depth analysis of the prevailing microbiota at the species level. The community members, isolated from the mature brew samples of brewery 1, were identified as *P. damnosus*, *A. pasteurianus* and *D. bruxellensis*, and their SSU rRNA gene sequences showed high similarities (> 97%) towards the pyrosequences of the respective most abundant OTUs. This species composition seems different from that reported by Martens and coworkers (1997), who mainly isolated *P. parvulus*, *D. lambicus*, *D. bruxellensis*, and some AAB (which were not identified further) at the end of the maturation phase.

P. damnosus is however phylogenetically closely related to *P. parvulus* and this species shares a lot of phenotypic characteristics with the latter (Holzapfel *et al.*, 2006). Martens and coworkers (1997) identified *P. parvulus* based on phenotypic and biochemical characteristics only, distinguishing *P. parvulus* from *P. damnosus* because of the inability of the latter to grow at 35°C. Yet, both species have been associated with the brewery environment or with environments in which alcoholic beverages are produced, with *P. damnosus* being the most frequently encountered in brewery environments (Sakamoto & Konings, 2003). Most likely, the fairly high concentrations of L-lactic acid and D-lactic acid found in the brew samples of the present study were associated with the presence of *P. damnosus*, which is able to metabolize glucose homofermentatively to both L- and D-lactic acid, and which creates a desired tartness. The total concentrations of lactic acid in the bottled beers measured in the present study were lower compared to those reported by Martens and coworkers (1997); these researchers also found a higher concentration of D-lactic acid compared to the concentration of L-lactic acid in the brew samples and bottled beers. Similarly to *P. damnosus*, *P. parvulus* produces both L-lactic acid and D-lactic acid but, to our knowledge, there is no information on the ratio of L-lactic acid to D-lactic acid produced by these species in the beer environment. At present, the reason for the observed difference remains unclear, although it is not excluded that the higher L-lactic acid concentration originates from acidification in the brewing process.

A. pasteurianus was isolated from a mature brew sample, which agrees with the occurrence of AAB such as *A. aceti*, *A. pasteurianus*, *A. lambicus*, and *Gluconobacter oxydans* (Bokulich & Bamforth, 2013; Spitaels *et al.*, 2014a) at the air-beer inter-

face and possibly in the porous wooden walls of the barrels (Bokulich *et al.*, 2012a; Martens *et al.*, 1997), and which are responsible for a change in beer flavor through the oxidation of ethanol into acetate. The high acetic acid concentrations detected in the present study are likely associated with the action of *A. pasteurianus*. Martens and coworkers (1997) found similar acetic acid concentrations in the mature brew samples, but they reported higher concentrations in the bottled beers compared to the findings of the present study. However, the blending protocol of this brewery has changed over time from a strongly acidic beer to a less sour alcoholic beverage (less lactic acid and acetic acid), following consumers' preferences.

Furthermore, taxonomic studies revealed that *D. lambicus* is a synonym of *D. bruxellensis* (Smith *et al.*, 1990), which confirms that *D. bruxellensis* is the dominant fungal member of this Belgian red acidic ale. *Dekkera bruxellensis* outcompetes *S. cerevisiae* in environments with low carbohydrate concentrations (Tiukova *et al.*, 2013). Metabolite target analysis showed that glucose was completely depleted at the end of the maturation phase, whereas substantial amounts of maltose and several oligosaccharides (maltotriose, -tetraose, -pentaose, and -hexaose) were still present. This contrasted with the results obtained for bottled beers, where the glucose and fructose concentrations were much higher and indicated that the end-product was sweetened by blending and possibly carbohydrate addition to comply with the consumers' preferences. *Saccharomyces cerevisiae* produces, along with ethanol and carbon dioxide, low molecular-mass flavor compounds, such as higher alcohols, aldehydes, ketones, organic acids, and esters that soften the sour taste and add fruity notes to the beers (Verstrepen *et al.*, 2003) as well as organic sulfides (Lodolo *et al.*, 2008). In the case of *D. bruxellensis*, it is known that this yeast is indispensable for the typical lambic beer flavor and that it produces several metabolites that are characteristic for lambic beers (Verachtert & Iserentant, 1995; Shanta Kumara & Verachtert, 1991; Spaepen *et al.*, 1978,9; Spaepen & Verachtert, 1982; Van Oevelen *et al.*, 1976). Metabolite target analysis of brew samples at the end of the maturation phase showed the presence of isoamyl alcohol and isoamyl acetate, which contrasted with the absence of 2-phenyl ethanol and 2-phenylethyl acetate. These metabolites were also found in the bottled beers. Propionic acid and isobutyric acid levels were low, as was described previously by Van Oevelen and colleagues (1976). The ethyl acetate (*i.e.*, solvent or fruity flavor) levels found in the brew samples at the end of the maturation phase and in the bottled beers of the present study were higher compared to what has been measured previously (Martens *et al.*, 1997). Whereas small concentrations of ethyl hexanoate and ethyl octanoate were found during the present study, no ethyl decanoate (*i.e.*, the typical ester present in lambic and gueuze beers (Van Oevelen *et al.*, 1976)) could be found. Next to the acetate esters mentioned above, these ethyl esters are also

flavor-active esters (Pires *et al.*, 2014), contributing to the taste of Belgian red acidic ales and gueuze beers (Spaepen & Verachtert, 1982).

Finally, no *Lactobacillus*, *Candida*, or other fungal species could be isolated from the mature brew samples in the present study, which was in agreement with previous findings (Martens *et al.*, 1997). Possible reasons are that these microorganisms are not able to grow under the experimental cultivation conditions, that they are in a viable but not cultivable state, or that they are no longer viable. Together, the findings on the microbial diversity and metabolite composition of Belgian red acid ales, obtained through 454 pyrosequencing and metabolite target analysis, show many similarities with those previously generated by Martens and coworkers (1997) with classic approaches.

The microbial diversity and metabolite composition of the brew samples from brewery 3 (which produced a red-brown acidic ale) resembled those of brewery 1 (which produced a red acidic ale), while brew samples from brewery 2 (which also produced a red-brown acidic ale) were very different. The microbial communities in brew samples from brewery 3 not only contained *Pediococcus* and *Dekkera*, but also an unclassified fungus and *Lactobacillus* (albeit in a lower concentration). In addition, lower relative abundances of the *Acetobacteraceae* and the *Candida* community members were also found in brew samples from brewery 3, which further contrasted with the findings in brew samples from brewery 1. The higher L-lactic acid and D-lactic acid and the lower acetic acid concentrations in brew samples from brewery 3 compared to those from brewery 1 could be attributed to these differences in community structures. It is not clear if these differences in relative abundances reflect the real structure of the microbial communities of the brew samples, because those abundances based on cultivation and molecular PCR-based identification could be partially attributed to amplification biases (Engelbrektson *et al.*, 2010). Additionally, several fungal OTUs remained unclassified in the brew samples from brewery 3. Different reference databases yielded different taxonomic assignments for these OTUs as a function of completeness and quality, underlining the need for a comprehensive, carefully annotated, fungal reference database (Begerow *et al.*, 2010; Seifert, 2009). Finally, both the microbial diversities and metabolite compositions of the three subsequent brews of brewery 2 were different from each other and from those of breweries 1 and 3. The high lactic acid and acetic acid concentrations of brew sample 2A in combination with a low number of AAB indicated that another community member, such as *Dekkera*, must have been responsible for the acetic acid production. The detection of minor fractions of the reads that were assigned to the genera *Weissella* and *Leuconostoc* in brew

sample 2B is remarkable. These LAB have not yet been isolated from beer, but were recently detected in a barcoded amplicon sequencing study of ACA (Bokulich *et al.*, 2012a).

The outcome of the multivariate analysis partially contrasted with what was known from literature. For instance, the presence of the *Pediococcus* community member was negatively correlated with L-lactic acid, whereas a positive correlation with isobutyl alcohol was found. This probably reflects the usage of lactate by other community members during fermentation. Furthermore, the *Acetobacteraceae* community member was not correlated with acetic acid, besides its positive correlation with isobutyl alcohol and negative correlation with L-lactic acid, contrasting with the metabolic features of this community member. Possibly, the current dataset is too small to unravel valid links between the microbiota and the metabolite compositions of mature Belgian red-brown acidic ales or that community members involved during fermentation and maturation collectively impact the flavor of these beers. Furthermore, the microbiota detected in the finished fermentations and beers might not necessarily represent the active populations during fermentation.

In addition to the high-abundance OTUs discussed above, numerous low-abundance OTUs were discovered during this study. The diverse taxonomical assignments included *Actinobacteria*, *Deinococcus-Thermus*, *Bacteroidetes*, *Beta-* and *Gammaproteobacteria*, *Kregervanrija*, *Debaryomyces*, *Priceomyces*, *Hyphopichia*, *Wickerhamomyces*, and others. Most of these microorganisms have not been reported in the brewery environment. A similar finding was seen in a study applying barcoded amplicon sequencing to unravel the microbiota of ACA (Bokulich *et al.*, 2012a). One possible explanation is that extensive sequencing of low-diversity microbial communities leads to the accumulation of errors that still remain present after stringent denoising (Dickie, 2010; Huse *et al.*, 2007; Kunin *et al.*, 2010). This leads to a long tail of low-abundant taxa in the species abundance curve, which are likely to be sequencing errors, and stresses the importance of focusing on the most abundant OTUs during amplicon sequencing studies that are performed without extensive biological and technical replicates. Another explanation may be that the physiological status of these microbes prevents their cultivation on the growth media used routinely. In that scenario, the present molecular community profiling method would be the first step to reveal these low-abundant microorganisms involved in Belgian red-brown acidic ale fermentations.

In conclusion, the microbial community diversity and metabolite composition of three Belgian red-brown acidic ales at the end of their maturation phase was

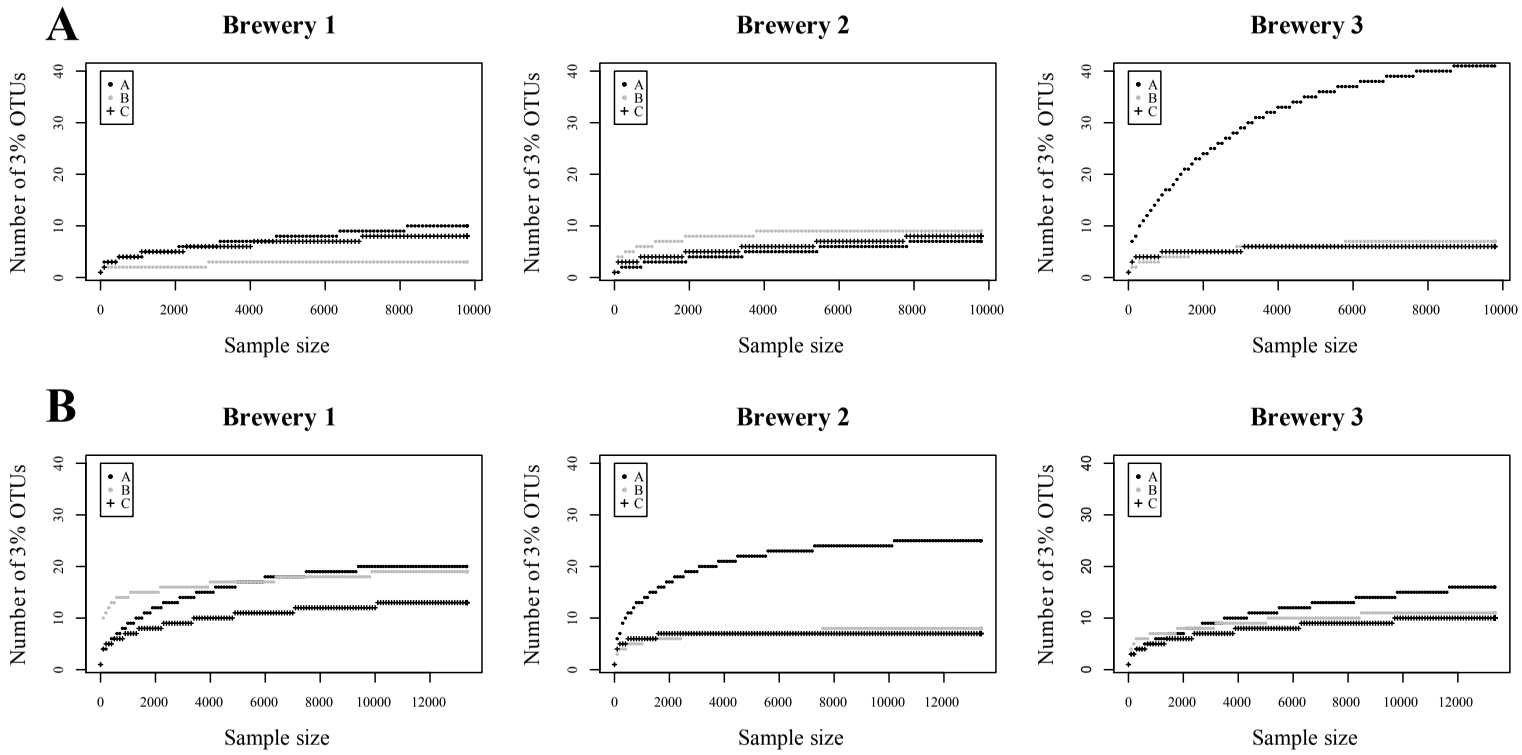
analyzed during the present study. In two of the brew samples of these breweries, the maturation phase was mainly dominated by *Pediococcus* and *Dekkera*. The microbial communities in brewery 2 samples where highly variable and differed from those of breweries 1 and 3. The main metabolites produced during these mixed-culture fermentations were ethanol, L- and D-lactic acid, and acetic acid. Ethyl acetate, isoamyl acetate, ethyl hexanoate, and ethyl octanoate dominated the flavor of the Belgian red-brown acidic ales of the present study.

5.1.5 Acknowledgments

The authors acknowledge the financial support of the FWO-Flanders, the Research Council of Ghent University (BOF project) and the Vrije Universiteit Brussel (SRP, IRP, and IOF projects), and the Hercules Foundation. In addition, the authors would like to acknowledge the breweries involved, the BCCM/LMG bacteria collection for providing DNA for MOCK community construction, Charlotte Peeters and Bart Verheyde for bioinformatics support, Leilei Li for sequencing the *groEL* housekeeping gene, and Tom Balzarini for the metabolite target analyses.

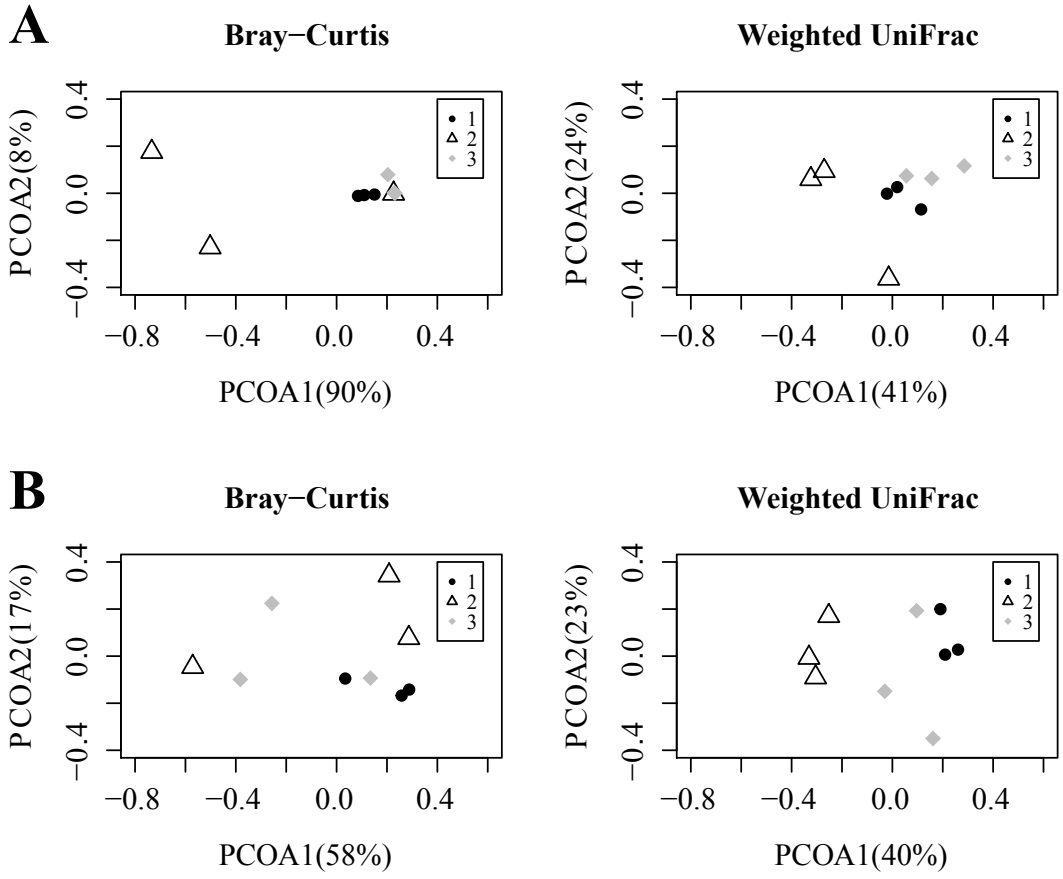
5.1.6 Supplementary Material

The raw sequence data received from Beckman Coulter were deposited at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) (accession number: SRP035529). The GenBank/EMBL accession number for the Sanger sequences generated during this study are LK024185-LK024190.



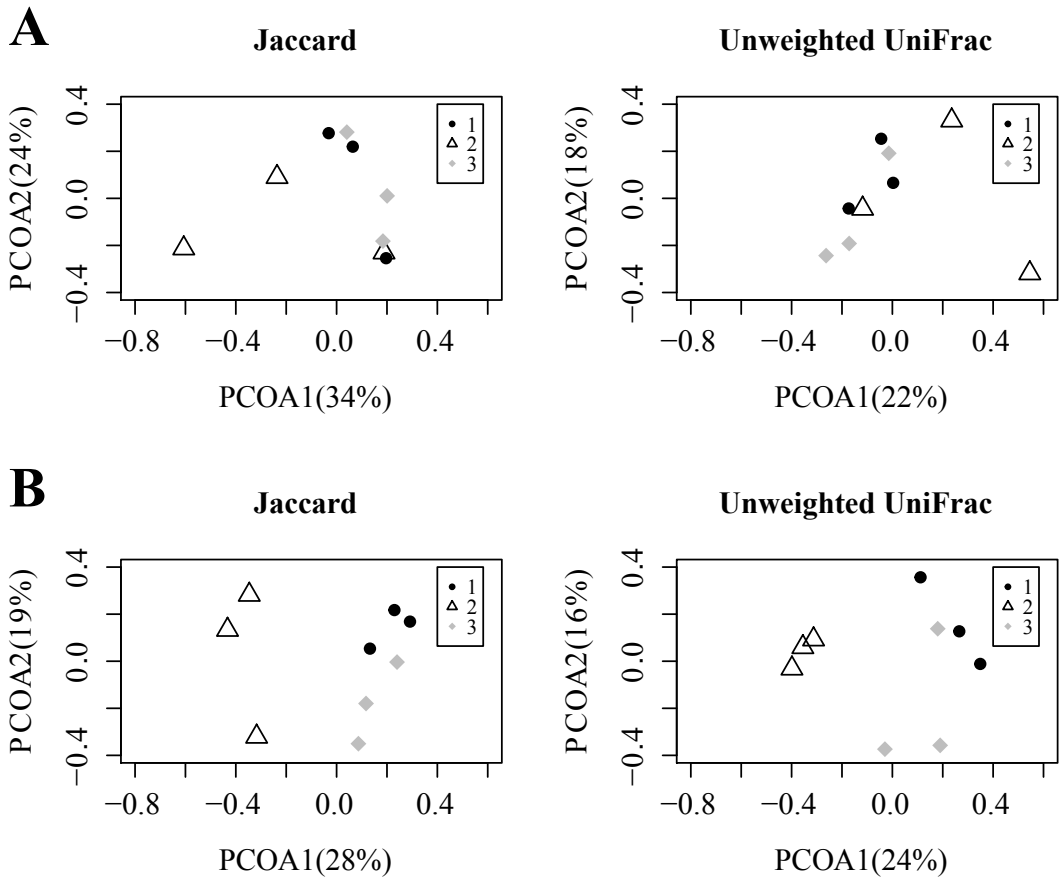
Supplementary Material 5.1: Rarefaction analysis of the normalized bacterial (A) and fungal (B) datasets of the three consecutive brew samples (A, B, and C) from breweries 1, 2, and 3

Number of 3% OTUs: *i.e.*, the richness of the OTUs at the 3% level.



Supplementary Material 5.2: PCoA plots of bacterial (A) and fungal (B) communities from brew samples that were subjected to 454 pyrosequencing

The positions of the bacterial and fungal communities for each species along the two first principal coordinate axes are illustrated, along with the percentage of variation explained by each axis. The results are based on the Bray-Curtis and weighted UniFrac distances, as indicated on the graphs.



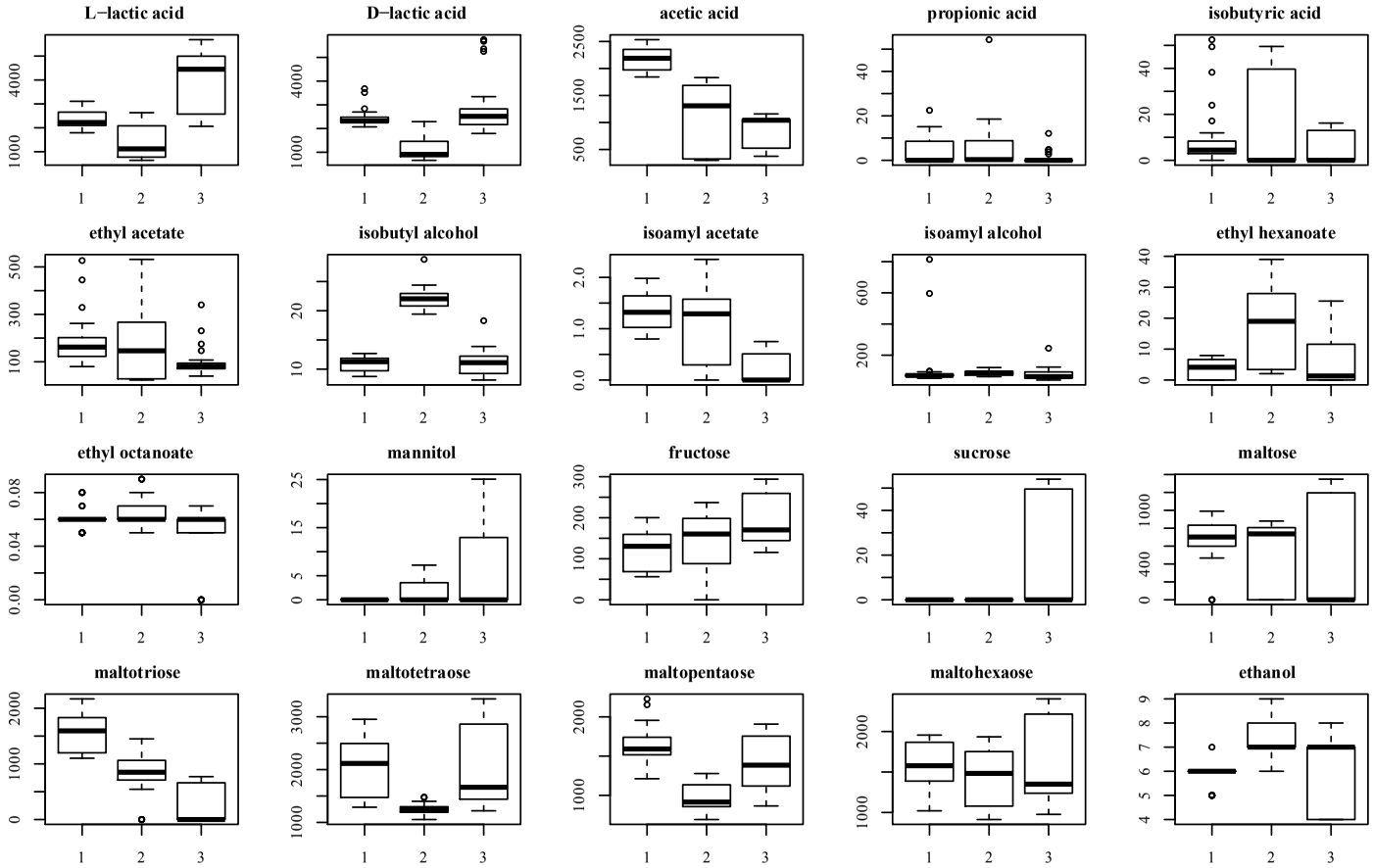
Supplementary Material 5.3: PCoA plots of bacterial (A) and fungal (B) communities from brew samples that were subjected to 454 pyrosequencing

The positions of the bacterial and fungal communities for each species along the two first principal coordinate axes are illustrated, along with the percentage of variation explained by each axis. The results are based on the Jaccard and unweighted UniFrac distances, as indicated on the graphs.

			Brewery 1			Brewery 2			Brewery 3		
			A	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
BACTERIAL DATASET	OTU 1	*Pediococcus*	86.51	88.23	83.87	99.44	1.87	23.27	90.27	95.11	92.77
	OTU 2	*Acetobacteraceae*	11.88	7.96	14.25	0.46	52.15	76.71	0.13	0.11	0.35
	OTU 3	*Lactobacillus*	0.01	0.01			33.44		3.31	0.04	0.09
	OTU 4	*Lactobacillus*	0.01	0.01			7.18		0.67	0.04	0.04
	OTU 5	*Lactobacillus*	1.45	3.71	1.32		0.01	0.02	5.37	4.66	6.70
	OTU 7	*Weissella*					2.02				0.01
	OTU 8	*Leuconostoc*					2.02				
FUNGAL DATASET	OTU 1	*Dekkera*	61.60	85.30	94.08	51.79	74.45	14.02	32.98	82.24	28.21
	OTU 2	unclassified	20.95	1.96	0.64	0.14	0.01	85.07	36.41	16.39	60.25
	OTU 3	*Candida*	15.00	11.00	4.96				0.01		9.45
	OTU 4	*Pichia*	0.01			37.45			0.31	0.08	0.03
	OTU 5	*Pichia*		0.10	0.19	9.93	11.40	0.02	10.42	0.89	0.28
	OTU 6	unclassified	0.01						6.88	0.26	
	OTU 7	*Candida*					11.08	0.01			
	OTU 8	unclassified							3.92		
	OTU 9	unclassified				0.01			3.19	0.10	
	OTU 10	*Kregervanrija*		1.03							1.15
	OTU 12	unclassified							1.78		
	OTU 13	unclassified							1.47		
	OTU 14	unclassified	1.98					0.01			
	OTU 15	unclassified							1.44		

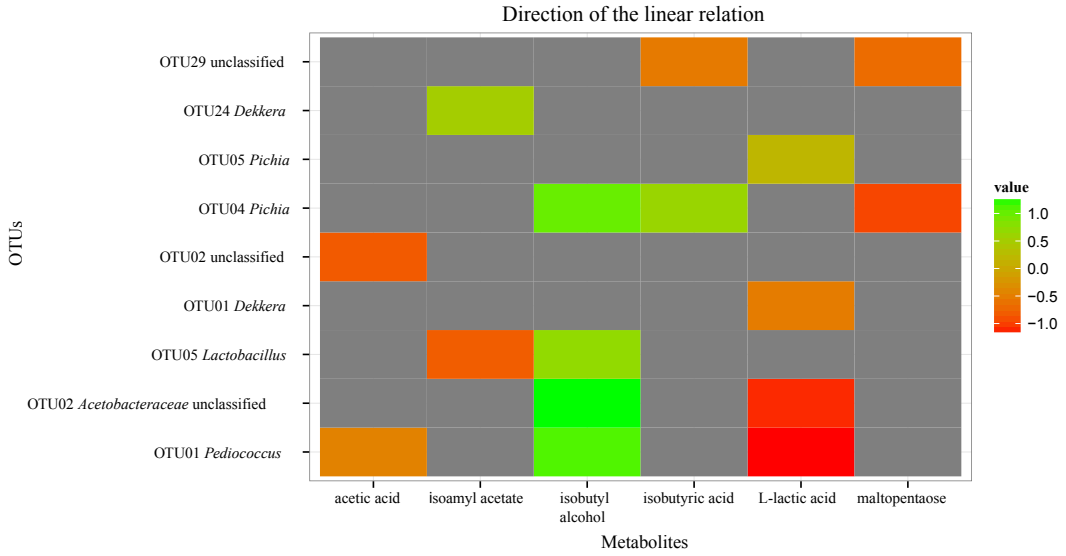
Supplementary Material 5.4: Relative abundances (%) of the most abundant (> 0.01% of the reads) bacterial and fungal OTUs used for the construction of the heatmap in Figure 15

Relative abundances of 0.00% are highlighted in grey.

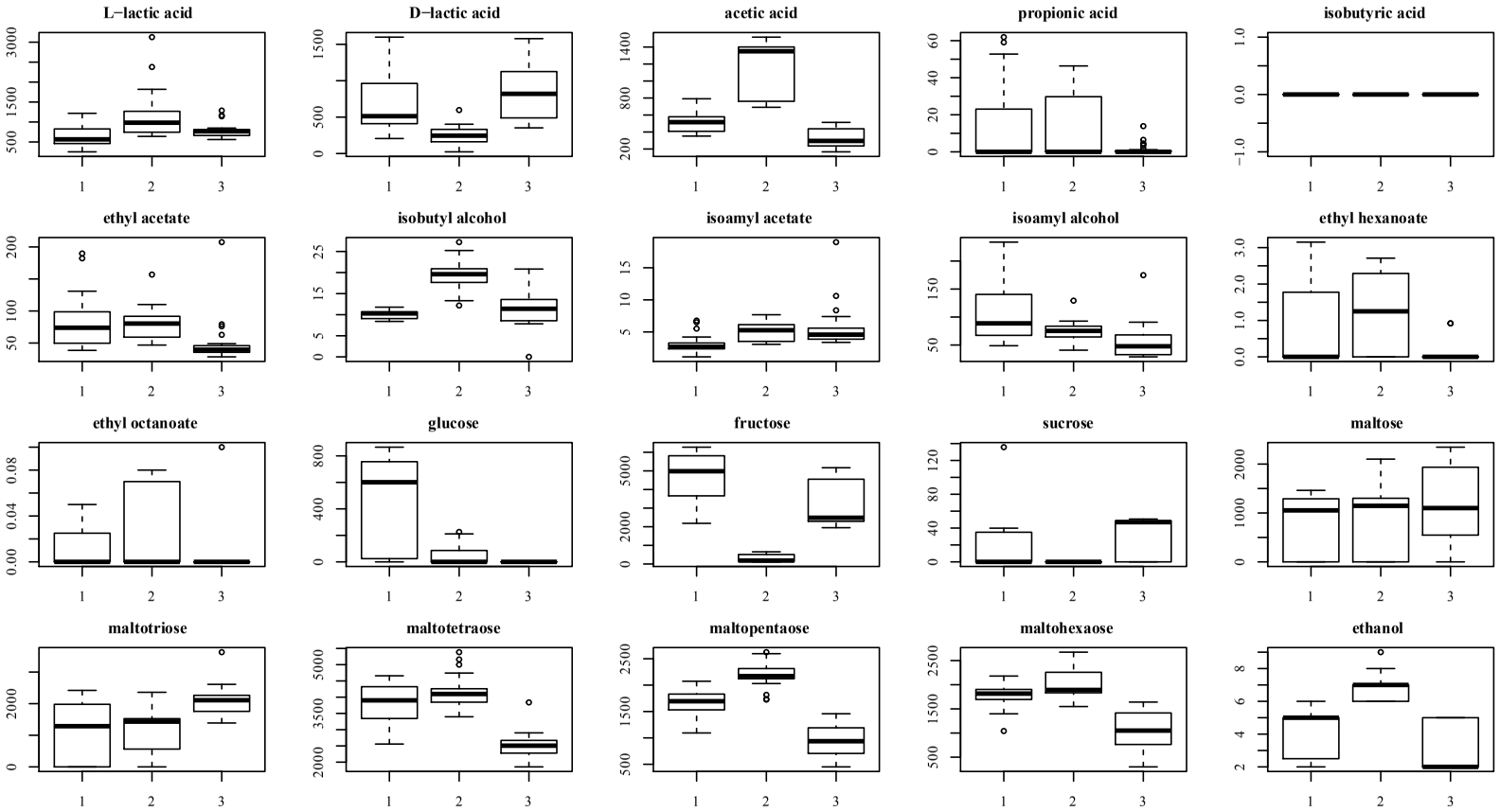


Supplementary Material 5.5: Boxplot of each of the metabolites measured in brew samples from the end of the maturation phase, in which the data points were grouped according to the brewery they originated from

The quantities are expressed in mg/L for all metabolites, except for ethanol [in (vol/vol)%]



Supplementary Material 5.6: Link between the microbial diversity and metabolite target analysis data set in brew samples from the end of the maturation phase
Highly positive and highly negative correlations are indicated in green and red, respectively. Correlations closer to zero are indicated in orange.



Supplementary Material 5.7: Boxplot of the metabolite concentrations in the bottled beers, in which the data points were grouped according to the brewery the beers originated from

The quantities are expressed in mg/L for all compounds, except for ethanol in (vol/vol)%.

5.1.7 Data Analysis Pipeline

The data analysis pipeline used is based on a paper of Schloss and coworkers (http://www.mothur.org/wiki/454_SOP) (2011a), with several modifications. Schloss and coworkers (2011a) constructed the pipeline based on the analysis of a mock community. The data analysis pipeline involves several steps, including (i) identifying and removing low quality sequences, (ii) trimming of sequences, (iii) denoising flowgrams, (iv) denoising sequences, (v) identifying and removing chimera's, etc.

Identifying and Removing Low-quality Sequences

Several features are linked with sequences of low quality. Therefore, identifying and removing sequences that contain these features reduces the overall error rate. Examples of features that are linked with low quality reads are the following: sequences with mismatches to the barcode of primer region, sequences that contain ambiguous base calls, sequences that are shorter than 200 bp, sequences that contain homopolymers, sequences that aligned to the incorrect region within the 16S rRNA gene... Similar parameters as applied by Schloss and coworkers (2011a) were used, with the exception of the number of mismatches to the barcode and primer region (which was set to zero) and the number of homopolymers (which was set to 6), for which more stringent values were chosen.

Trimming of Sequences

The association between error rates and quality scores, as shown by Schloss and coworkers (2011a), indicated that trimming sequences to a break point, where quality score criteria were no longer met, results in reducing the overall error rates. Schloss and coworkers (2011a) proved that a 50 bp sliding window with an average quality score of 35 within the window was the most efficient. Exactly the same approach was used in the present study. Furthermore, the sequences were trimmed to a region were all sequences started and ended in the same alignment positions. This trimming insured that evolutionary consistent regions were being compared. For this purpose, our data was aligned to the SILVA-compatible template alignment, containing 50000 columns. The general approach was (i) to find the closest template for each candidate using kmer (k=8) searching, (ii) to make a pairwise alignment between the candidate and de-gapped template sequences using the Needleman-Wunsch algorithm with a reward of +1 for a match and penalties of -1 and -2

for a mismatch and gap, respectively and (iii) to re-insert gaps to the candidate and template pairwise alignments using the NAST algorithm so that the candidate sequence alignment is compatible with the original template alignment. The aligner does not explicitly take into account the secondary structure of the 16S rRNA gene, but the SILVA-compatible template alignment is based on the secondary structure, resulting in an alignment that will at least be implicitly based on the secondary structure.

Denoising Flowgrams

The Pyronoise algorithm (Quince *et al.*, 2011) was applied, which reduces the sequencing error rate by correcting the original flowgram data using an expectation-maximization algorithm. This algorithm was re-implemented in the mothur software package to take advantage of accelerated clustering algorithms and to make the algorithm more accessible to other researchers. Flowgrams were trimmed to 450 flows because of the reasons explained by Schloss and coworkers (2011a).

Denoising Sequences

Lingering PCR amplification and sequencing errors were removed using the SeqNoise algorithm. This algorithm uses the sequences as inputs and a model that describes rates of substitutions and homopolymeric insertions and deletions.

Identifying and Removing Chimera's

Schloss and coworkers explored several chimera removal strategies, *e.g.*, Chimera Slayer, Uchime, and Perseus. They showed that Uchime and Perseus are preferred above the Chimera Slayer software. We chose to identify and remove chimera's using the Uchime software. As shown by Schloss and coworkers, Uchime is preferred to Perseus because of its lower false discovery rate and because of the previously reported faster execution times.

5.2 Comparative Genome Analysis of *Pediococcus damnosus* LMG 28219, a Strain Well-Adapted to the Beer Environment

I. Snauwaert, P. Stragier, L. De Vuyst, P. Vandamme

Submitted to: BCM Genomics, 2014

Author contributions: IS carried out DNA extraction, conducted genome assembly, performed sequence annotation, bioinformatics analyses, and comparative genome analyses. PS performed the OrthoMCL analysis. IS and PV designed and coordinated the study and participated in the analysis of the results. IS, PS, LDV, and PV wrote the manuscript. All authors read and approved the final manuscript.

Pediococcus damnosus **LMG 28219 is a LAB dominating the maturation phase of Flemish acid beer productions. It proved to be capable of growing in beer, thereby resisting this environment, which is unfavorable for microbial growth. The molecular mechanisms underlying its metabolic capabilities and niche adaptations are currently unknown. In the present study, whole-genome sequencing and comparative genome analysis were used to investigate this strain's mechanisms to reside in the beer niche. The draft genome sequence of *P. damnosus* LMG 28219 harbored 183 contigs, including an intact prophage region and several CDSs involved in plasmid replication. The annotation of 2178 CDSs revealed the presence of many transporters and transcriptional regulators and several genes involved in oxidative stress response, hop resistance, *de novo* folate biosynthesis, and EPS production. Comparative genome analysis of *P. damnosus* LMG 28219 with *Pediococcus claussenii* ATCC BAA-344^T (beer origin) and *Pediococcus pentosaceus* ATCC 25745 (plant origin) revealed that various hop resistance genes and genes involved in *de novo* folate biosynthesis were unique to the strains isolated from beer. This contrasted with the genes related to osmotic stress responses, which were shared between the strains compared. Furthermore, transcriptional regulators were enriched in the genomes of bacteria capable of growth in beer, suggesting that those cause rapid up- or down-regulation of gene expression. Genome sequence analysis of *P. damnosus* LMG 28219 provided insights into the underlying mechanisms of its adaptation to the beer niche. The results presented will enable analysis of the transcriptome and proteome of *P. damnosus* LMG 28219, which will result in additional knowledge on its metabolic activities.**

5.2.1 Introduction

Beer is a fermented beverage that is high in ethanol, carbon dioxide and flavorful yeast metabolites, contains hop-derived flavor and antimicrobial components, and is low in pH, oxygen, and residual nutrients (Bokulich & Bamforth, 2013). This environment has selected for unique groups of bacteria specialized in growth in beer, including several species of LAB. Overall, one of the key metabolic actions of LAB is the reduction of pyruvate into lactate to regenerate NAD^+ by means of lactate dehydrogenase activity. Depending on the beer type produced, lactate production may or may not be desired (Bokulich *et al.*, 2012a; Martens *et al.*, 1997; Van Oevelen *et al.*, 1977; Spitaels *et al.*, 2014b). As most LAB species tolerate high ethanol concentrations and a low pH, hop resistance is the major factor limiting their growth in beer. Hop-derived antimicrobial compounds, known as iso-alpha acids, cause permeability changes in the bacterial cell wall (Shimwell, 1937), leakage of the cytoplasmic membrane and subsequent inhibition of respiration and protein, DNA, and RNA syntheses (Teuber & Schmalreck, 1973), as well as changes in leucine uptake and proton ionophore activity (Simpson, 1993). Additionally, iso-alpha acids alter the redox properties of the bacterial cell, causing oxidative stress (Behr & Vogel, 2010). Accordingly, hop resistance is a multifactorial property that implies different mechanisms to counteract the action of iso-alpha acids. A key factor mediating hop resistance is the ATP-binding cassette multidrug transporter, HorA, which extrudes iso-alpha acids (Sakamoto *et al.*, 2001) and is plasmid-encoded (Sakamoto & Konings, 2003).

LAB species frequently encountered in the beer environment belong to the genera *Pediococcus* and *Lactobacillus*, which are both Gram-positive, facultative anaerobic, chemo-organotrophic bacteria of the family *Lactobacillaceae* of the *Firmicutes* phylum. The species with the highest capacity to grow in beer are *Lactobacillus brevis* and *Pediococcus damnosus*, although the ability of bacteria to grow in beer is a strain- rather than species-specific characteristic (Pittet *et al.*, 2011). Recent studies investigated the mechanisms of *Lactobacillus brevis* for overcoming stresses in beer by means of reverse transcription qPCR and proteomic analyses (Behr *et al.*, 2007; Bergsveinson *et al.*, 2012). Less is known about the adaptations of *P. damnosus* to the beer environment. So far, *P. claussenii* ATCC BAA-344^T originating from spoiled beer, *P. pentosaceus* ATCC 25745 of plant origin, and *P. pentosaceus* IE-3 originating from a dairy effluent sample, are the only members of the genus *Pediococcus* for which the genome has been sequenced completely

(Midha *et al.*, 2012; Pittet *et al.*, 2012; Makarova *et al.*, 2006). However, draft genome sequences are available for several strains of *Pediococcus acidilactici*, namely strain MA 18/5M originating from animal food, strain DSM 20284 isolated from barley, strain 7_4 from a human fecal sample, and strain NGRI 0510Q, formerly known as *Pediococcus lolii* (Wieme *et al.*, 2012), isolated from ryegrass silage (`http://genomesonline.org`). In general, *Pediococcus* species possess rather small genomes (approximately 2 MB), encoding a broad repertoire of transporters for efficient carbon and nitrogen acquisition and reflecting a limited range of biosynthetic capabilities. This suggests both extensive gene loss as well as acquisitions via horizontal gene transfer during the evolution of pediococci within their habitats (Makarova *et al.*, 2006). Several strains have plasmids containing genes regulating the fermentation of carbohydrates and encoding different types of resistances.

The strain *P. damnosus* LMG 28219 was isolated in 2013 from a Flemish acid beer at the end of its maturation phase. Flemish acid beers are produced by a mixed-culture fermentation and represent culturally important products, for which microbial activities play critical roles in beer production and quality formation (Martens *et al.*, 1997; Spitaels *et al.*, 2014b). In the present study, the draft genome sequence of *P. damnosus* LMG 28219 is presented and analyzed to obtain insights into its genome-based metabolic features. A better understanding of the molecular mechanisms underlying its metabolic capabilities enabled detailed insights into the mechanisms of adaptation of this strain to the beer environment. Furthermore, comparison of *P. damnosus* LMG 28219 with other sequenced members of the genus *Pediococcus* addressed the potentially unique properties of this strain and other strains adapted to the beer environment.

5.2.2 Materials and Methods

Bacterial Strain and Growth Conditions

The strain *P. damnosus* LMG 28219 was used for genome sequencing. It was identified through sequence analysis of the *pheS* gene, as described previously (De Bruyne *et al.*, 2008b). To obtain cell pellets, *P. damnosus* LMG 28219 was propagated in MRS broth (Oxoid, Basingstoke, Hampshire, United Kingdom) at 28°C for 3 days, followed by microcentrifugation (11000 rpm, 15 min, 4°C).

DNA Extraction and Illumina Sequencing

Total DNA was extracted using a procedure applied by Gevers and colleagues (2001), with several modifications. (i) The thawed pellet was washed in 1 ml TES buffer (6.7% sucrose, 50 mM Tris-HCl, pH 8.0, 1 mM EDTA) and resuspended in 275 μL STET buffer (8.0% sucrose, 5.0% Triton X-100, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA). (ii) Ninety μL lysozyme-mutanolysin-proteinase K solution (TES buffer containing 1667 U/mL mutanolysine, 33 mg/mL lysozyme and 2.78 mg/mL proteinase K) was added and the suspension was incubated at 37°C for 1 h. (iii) Prior to extraction with a phenol/chloroform/isoamylalcohol (49.5:49.5:1.0) solution, proteins were precipitated by adding 250 μL ammonium acetate. (iv) Five μL RNase (2 mg/L) was added to the DNA solution, which was incubated at 37°C for 60 min and stored at -20°C. The integrity, concentration, and purity of the DNA isolated was evaluated using 1.0% (wt/vol) agarose gels, stained in ethidium bromide, and by spectrophotometric measurements at 234, 260, and 280 nm. A fluorescent stain-based kit (Qubit® dsDNA Broad range assay kit; Life Technologies, Carlsbad, CA, USA) was used for an accurate determination of the DNA concentration. Because of the low DNA concentration obtained, the extraction procedure was performed in triplicate and the resulting DNAs were pooled, after which an additional purification [by adding 12.5 μL sodium acetate (3 M, pH 4.8) and 200 μL 95% ice-cold ethanol to 100 μL DNA product] and quantification step was performed.

Illumina Sequencing and Genome Assembly

Library preparation and genome sequencing was performed by BaseClear BV (Leiden, The Netherlands). A paired-end DNA library with a mean insert length size between 230 and 360 bp was sequenced, with average reads of 100 bp, on an Illumina HiSeq2500 apparatus (Illumina Inc., San Diego, CA, USA). An initial *de novo* assembly was performed in CLC Genomics Workbench v6.5.1 (CLC Inc., Aarhus, Denmark), using quality-trimmed (based on a threshold of Q = 30) and paired reads. All contigs shorter than 500 bp were discarded. Contigs were ordered automatically with MAUVE (Darling *et al.*, 2004) by comparing the draft genome sequence of *P. damnosus* LMG 28219 with the complete chromosomal DNA of *P. claussenii* ATCC BAA-344^T. Contigs that did not show similarities towards the *P. claussenii* ATCC BAA-344^T chromosomal DNA were blasted against a plasmid and phage database, using the PATRIC (*i.e.*, Pathosystems Resource Integration Center) (Wattam *et al.*, 2014) and PHAST (*i.e.*, PHAge Search Tool) (You *et al.*, 2011) websites, respectively. In addition, the average G+C content of the contigs was calculated using an in house developed Python script.

Genome Annotation

Functional annotation and metabolic reconstruction were performed with (i) the RAST server (Aziz *et al.*, 2008), using GLIMMER (Delcher *et al.*, 1999) for gene calling and allowing frameshift corrections, backfilling of gaps, and automatic fixing of errors; (ii) the IMG-Expert review (IMG-ER) annotation pipeline, using GenePRIMP for gene prediction (Mavromatis *et al.*, 2009); and (iii) the NCBI's PGAAP, which uses GeneMark and GeneMark.HMM for gene calling (Lukashin & Borodovsky, 1998). The automated gene prediction and annotation was followed by manual curation of the CDSs of interest using BLASTp (Altschul *et al.*, 1997), UniProt (`http://www.uniprot.org/`), and InterProScan (Zdobnov & Apweiler, 2001).

Comparative Genomics

Comparative genomics of the draft genome sequence of *P. damnosus* LMG 28219 was performed using the available complete genome sequences of *P. claussenii* ATCC BAA-344^T and *P. pentosaceus* ATCC 25745. The ANIs between the draft genome (*P. damnosus* LMG 28219) and both reference genomes were calculated using the *in silico* DDH method implemented in the Jspecies software (Rosselló-Mora, 2006), using BLAST, as proposed by Goris and colleagues (2007). The OrthoMCL tool was used with default settings for ortholog finding in the genomes of *P. damnosus* LMG 28219, *P. claussenii* ATCC BAA-344^T, and *P. pentosaceus* ATCC 25745 (`http://orthomcl.org/`). The Kyoto encyclopedia of genes and genomes (KEGG) database was used for the reconstruction of the folate biosynthesis metabolic map (`http://www.genome.jp/kegg/`).

5.2.3 Results and Discussion

General Architecture and Annotation of the *Pediococcus damnosus* LMG 28219 Draft Genome

Paired-end sequencing of the *P. damnosus* LMG 28219 genomic DNA yielded 3,137,316 reads with a total number of 2,231,216 bp that were assembled into 183 contigs (N50 of 24659), consisting of 69 large (> 10,000 nucleotides) and 114 small

(< 10,000 nucleotides) contigs. This genome hence represents an intermediate size among the LAB (Midha *et al.*, 2012; Makarova *et al.*, 2006). An overview of the G+C content and the length of the contigs is presented in Supplementary Material 5.8. The G+C content of the complete draft genome averaged 38.2 mol%. A total of 91 contigs could be mapped onto the *P. claussenii* ATCC BAA-344^T chromosomal DNA and were ordered accordingly (Table 8). Two clustered regularly interspaced short palindromic repeat (CRISPR) arrays were found on contigs 71 and 150, whereas three CRISPR-associated CDSs (AH70_09625, AH70_09630, and AH70_09635) were found on contig 71. One intact prophage region (G+C content, 39.3%; region length, 39.4 kb) containing 50 CDSs was predicted and identified as the *Lactobacillus* phage Sha1, which was originally isolated from kimchi (Yoon *et al.*, 2011). This phage has an isomeric head and a long tail and is classified as a member of the large family of *Siphoviridae*. The genes of phage Sha1 are organized into five functional clusters: replication/regulation/modification, packaging, structure/morphogenesis, lysis, and lysogeny. Most of the phage-related CDSs were found on contig 7, which was predicted to be part of the chromosomal DNA.

A BLAST search of the remaining non-chromosomally encoded contigs against a plasmid-specific database revealed many similarities towards plasmids of different bacterial species (see below). Many of these contigs (as listed in Table 8) were similar to the plasmids of *P. claussenii* ATCC BAA-344^T, with the exception of plasmids pPECL-1 and pPECL-2, onto which no contigs mapped. These two plasmids are small and cryptic, whereas the other six plasmids of *P. claussenii* ATCC BAA-344^T range from 16 to 36 kb and contribute to roughly 7% of the strain's coding capacity (Pittet *et al.*, 2013). Four, six, ten, four, eight, and three contigs mapped onto plasmids pPECL-3, pPECL-4, pPECL-5, pPECL-6, pPECL-7, and pPECL-8 of *P. claussenii* ATCC BAA-344^T, respectively. A total of 19 contigs did not show high similarities towards the plasmids of *P. claussenii* ATCC BAA-344^T. Contigs 16 and 57 were similar to plasmids pBM400 and WSH-002_p1 of *Bacillus megaterium* strains QM B1551 and WSH-002, respectively, whereas others were similar to plasmids of *Lactobacillus brevis* strains 925A, KB290, and ATCC 367 (contigs 123, 134, and 140, respectively), *Lactobacillus plantarum* WCFS1 (contig 37), *Lactobacillus buchneri* strains CD034 and NRRL B-30929 (contigs 100 and 122, respectively), or other *Lactobacillus* species (all remaining contigs). In addition, several CDSs involved in plasmid replication were found, which encoded replication initiation proteins (AH70_10315 on contig 8, AH70_03600 on contig 21, AH70_00575 on contig 107, AH70_02015 on contig 138, and AH70_03140 on contig 182).

Table 8: Overview of the P. damnosus *LMG 28219 draft genome.*

Draft genome of *P. damnosus* LMG 28219	Mean G+C content (%)	Mean size consensus (bp)	Best BLAST hit
82, 98, 145, 116, 77, 119, 47, 53, 5, 89, 56, 76, 166, 84, 81, 136, 102, 141, 72, 30, 52, 129, 74, 54, 19, 110, 147, 49, 26, 85, 137, 61, 146, 177, 7, 68, 131, 104, 87, 32, 17, 80, 66, 165, 2, 20, 101, 96, 4, 59, 45, 164, 118, 64, 176, 69, 90, 24, 127, 10, 46, 139, 114, 60, 124, 23, 83, 51, 151, 169, 106, 11, 121, 78, 55, 43, 130, 48, 65, 168, 27, 62, 60, 79, 6, 152, 120, 73, 35, 99, 29, 71*	38.3	20814	chromosomal DNA (NC_016605)
21, 39, 41, 183	36.9	4040	pPECL-3 (NC_016636)
12, 21, 94, 95, 112, 135	38.5	3419	pPECL-4 (NC_016607)
1,21, 39, 40, 95, 97, 107, 132, 135, 170	38.6	5198	pPECL-5 (NC_016608)
21, 107, 135, 178	37.0	4924	pPECL-6 (NC_017017)
8,21, 39, 41, 86, 94, **182**, 183	38.1	3926	pPECL-7 (NC_017018)
70,**14**, **97***	42.5	9516	pPECL-8 (NC_017019)
16, 57	49.0	2538	*Bacillus megaterium* QM B1551 (pBM400, NC_004604) and WSH-002 (WSH-002_p1, CP003018)
123, 134, 140	37.2	4034	*Lb. brevis* 925A (pLB925A04, NC_012551), KB290 (pKB290-4, AP012171), and ATCC 367 (plasmid 1, NC_008498)
37	44.6	1242	*Lb. plantarum* WCFS1 (pWCFS103, NC_006377)
100, 122	33.9	5879	*Lb. buchneri* CD034 (pCD034-3, CP003044) and NRRL B-30929 (pLBUC01, NC_015420)
18, 22, 93, 108, 109, 113, 115, 143, 158, 162, 163	40.6	2797	Several *Lactobacillus* spp.

Contigs with the highest hit scores are highlighted in bold. *Contig numbers are mentioned in ordered fashion. p: plasmid, *Lb.*: *Lactobacillus*.

Gene finding and annotation of the *P. damnosus* LMG 28219 draft genome with the IMG-ER software resulted in a total gene count of 2266, among which 96.12% were CDSs. A total of 79.08% of the CDSs could be assigned to a protein, among which 23.39% and 12.80% were predicted to be enzymes and transporters, respectively. This agrees with the property of LAB to encode a broad repertoire of transporters for efficient nutrient uptake and reflects their limited biosynthetic capabilities (Makarova *et al.*, 2006). Additionally, 2.60% of the genes encoded signal peptides, whereas 26.04% encoded transmembrane proteins. Furthermore, a total of 56 tRNA and three 5S, 16S, and 23S rRNA genes were predicted. As expected, most CDSs were involved in carbohydrate, protein, DNA, and RNA metabolism and in cell wall and capsule construction. These findings agreed with those of Makarova and colleagues (2006), who analyzed 12 genomes belonging to the order *Lactobacillales* and defined a core of *Lactobacillales*-specific clusters of orthologous groups (LaCOGs). The functional distribution of the conserved core of 567 LaCOGs showed that the majority encodes components of the information processing systems (translation, transcription, and replication), which are likely to perform essential functions. With the exception of 9 LaCOGs (*i.e.*, COGs 0230, 0419, 0454, 0457, 0470, 0762, 2815, 2855, and 4608), all core LaCOGs were found in the draft genome of *P. damnosus* LMG 28219. Genome closure could possibly lead to the identification of these missing core functions, although an update of the core LaCOGs is mandatory, because only 12 *Lactobacillales* species out of more than 500 species of this order were included; in addition, only one of these 12 species belonged to the genus *Pediococcus*.

Next to these core functions, the draft genome sequence of *P. damnosus* LMG 28219 comprised 50 predicted CDSs that were associated with virulence, disease, and defense, including three multidrug resistance efflux pumps. Also, 42 CDSs related to stress response were detected, seven of which were involved in oxidative stress response. Four of the non-core LaCOGs present in the *P. damnosus* LMG 28219 draft genome comprised 12 or more CDSs in a COG, namely arabinose efflux permeases (COG2814), transcriptional regulators (COG1309), predicted transcriptional regulators (COG0789), and transposases and their inactivated derivatives (COG2826). Arabinose efflux permeases belong to the major facilitator superfamily, a large and diverse group of secondary transporters that includes uniporters, symporters, and antiporters (Reddy *et al.*, 2012). A total of 12 and 10 CDSs of COG2814 were annotated as arabinose efflux permeases and drug resistance transporters from the EmrB/QacA subfamily, respectively. In addition, one CDS was predicted to be a drug resistance transporter of the Bcr/CflA subfamily and one was predicted to be a transporter belonging to the sugar transporter family. The COG1309 members belonged to the TetR family of transcriptional repressors, which

are involved in transcriptional control of multidrug efflux pumps, pathways for the biosynthesis of antibiotics, responses to osmotic stress and toxic chemicals, control of catabolic pathways, differentiation processes, and pathogenicity. Analysis of the CDSs of COG0789 did not reveal insights into their function. Finally, the CDSs in COG2826 were all transposases belonging to the insertion sequence IS30 family. The presence of IS elements may lead to genomic instability due to its role in gene loss and gene gain (Darmon & Leach, 2014).

Comparative Genomics

The draft genome sequence of *P. damnosus* LMG 28219 was compared with the complete genome sequences of *P. pentosaceus* ATCC 25745 and *P. claussenii* ATCC BAA-344^T. The number of CDSs predicted in the draft genome of *P. damnosus* LMG 28219 was higher compared to those in the reference genomes (Table 9). In addition, the number of genes in *P. claussenii* ATCC BAA-344^T exceeded that of *P. pentosaceus* ATCC 25745. This could be explained by the need to acquire additional functions during evolution to be able to survive in the hostile beer environment, compared to the plant environment. Possibly, *P. pentosaceus* ATCC 25745 can acquire more nutrients from its environment compared to *P. damnosus* LMG 28219 and *P. claussenii* ATCC BAA-344^T. The number of rRNA operons in the genome of *P. damnosus* LMG 28219 was lower compared to that in *P. pentosaceus* ATCC 25745 and *P. claussenii* ATCC BAA-344^T, which may reflect differences in the ecological competitiveness (Makarova *et al.*, 2006; Klappenbach *et al.*, 2000; Di Mattia *et al.*, 2002). In addition, the ANIs of *P. damnosus* LMG 28219 versus *P. pentosaceus* ATCC 25745 and *P. claussenii* ATCC BAA-344^T were 69.81% and 69.28%, respectively, whereas the ANIs between *P. pentosaceus* ATCC 25745 and *P. claussenii* ATCC BAA-344^T were 71.53%. These results agreed with previous findings, indicating that *P. claussenii* and *P. pentosaceus* are evolutionarily closer related to each other compared to *P. damnosus* (De Bruyne *et al.*, 2008b).

The predicted proteome of *P. damnosus* LMG 28219 was assigned into orthologous clusters, along with the proteomes of *P. claussenii* ATCC BAA-344^T and *P. pentosaceus* ATCC 25745 to predict unique and/or shared characteristics between these LAB species. A total of 1062 putative orthologous proteins were shared between *P. pentosaceus* ATCC 25745 (of plant origin), *P. claussenii* ATCC BAA-344^T, and *P. damnosus* LMG 28219 (both of beer origin), whereas 25, 30, and 30 orthologs were unique to *P. pentosaceus* ATCC 25745, *P. claussenii* ATCC BAA-344^T, and *P. damnosus* LMG 28219, respectively (Figure 18; Supplementary

Table 9: Comparison of genome characteristics of P. damnosus *LMG 28219 with* P. pentosaceus *ATCC 25745 and* P. claussenii *ATCC BAA-344^T*

	P. pentosaceus **ATCC 25745**	***P. claussenii*** **ATCC BAA-344^T**	***P. damnosus*** **LMG 28219**
Accession number*	CP000422	CP003137*	JANK00000000
Origin	plant	beer	beer
Size	1,832,387	1,966,362	2,231,216
CDSs	1755	1892	2178
rRNA genes	5	4	3
tRNA genes	55	57	56
GC%	37.4	37.0	38.2

* Only the Genbank record of the chromosomal DNA is given

Material 5.9). Genes that are shared between *P. damnosus* LMG 28219 and *P. claussenii* ATCC BAA-344^T, but not by *P. pentosaceus* ATCC 25745, possibly promote survival in beer. For instance, the presence of genes involved in hop resistance, osmotic stress response, EPS production, and the presence or absence of complete or partial metabolic pathways may lead to the strain's ability to grow in beer. Furthermore, the functions enriched in the genomes of bacteria capable of growing in beer are likely to contribute to their specific phenotypic traits.

Genes Involved in Hop Resistance

Research on marker genes associated with hop resistance suggested that these genes are typically acquired via horizontal gene transfer (Suzuki *et al.*, 2004; Iijima *et al.*, 2007; Hayashi *et al.*, 2001). Although several putative hop resistance genes have been described, only the presence of *horA* definitively correlates with LAB growth in beer, with the presence of *horA* together with *hitA* and/or *horC* allowing fast bacterial growth (Bergsveinson *et al.*, 2012). A recent transcriptome study of *P. claussenii* ATCC BAA-344^T showed increased transcript levels of several genes during growth in beer compared to growth in MRS broth (Pittet *et al.*, 2013). This was especially the case for genes encoded on plasmids pPECL-3, pPECL-5, and pPECL-8. Contrasting to pPECL-3 and pPECL-5, many CDSs on pPECL-8 showed strong homology towards contigs 70, 14, and 97 in the draft genome of *P. damnosus* LMG 28219 (Figure 19). Plasmid pPECL-8 harbored the previously mentioned *horA* gene, which is involved in hop resistance. Surprisingly, several orthologs of the *P. claussenii* ATCC BAA 344^T *horA* gene were found both in the draft genome of *P. damnosus* LMG 28219 (AH70_02075 on contig

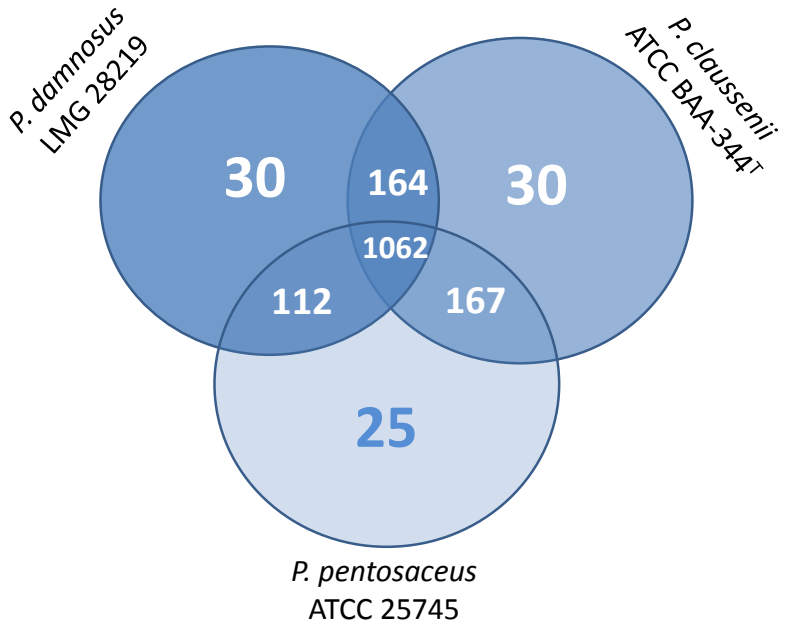


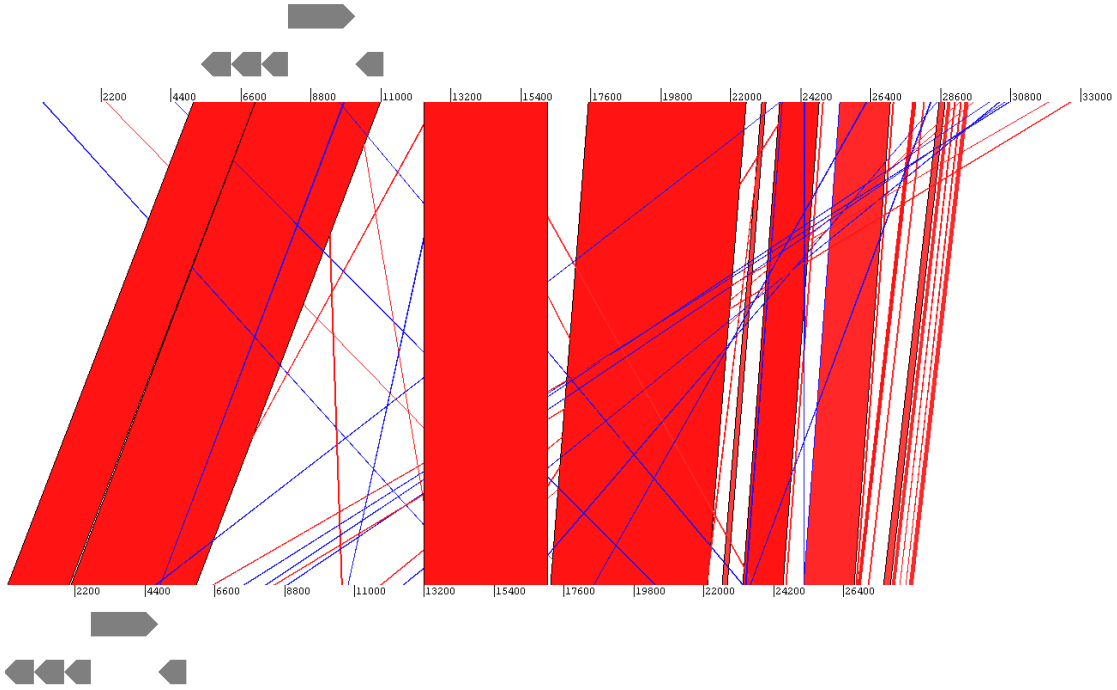
Figure 18: Visualization of the OrthoMCL output comparing the number of unique and/or shared orthologs of P. claussenii *ATCC BAA-344^T*, P. pentosaceus *ATCC 25745, and* P. damnosus *LMG 28219*

14 and AH70_01035 on contig 114) and in the complete genome of *P. pentosaceus* ATCC 25745 (YP_805121), which all encoded ABC-type multidrug transporters. In depth analysis of these CDSs revealed that only AH70_02075 showed high (99%) amino acid sequence homology towards the *horA* gene of *P. claussenii* ATCC BAA 344^T; the remaining CDSs showed less than 70% homology. Possibly, these *horA* orthologs are descendants of a common ancestor that evolved towards a different functionality, with the *horA* gene in *P. claussenii* ATCC BAA-344^T and AH70_02075 in *P. damnosus* LMG 28219 being specialized in the extrusion of iso-alpha acids.

Interestingly, several studies revealed that the *horA* gene is consistently surrounded by the same set of genes in bacteria growing in beer, suggesting that these entire regions were acquired via horizontal gene transfer (Pittet *et al.*, 2013; Iijima *et al.*, 2007) and could possibly play a role in their adaptation to the beer niche. Most of these *horA*-surrounding genes are involved in phospholipid or cell wall biosynthesis. Orthologs of PECL_1950, PECL_1952, and PECL_1954 flanking the *horA* gene were found in the genome of *P. damnosus* LMG 28219 (AH70_09615 on contig 70 and AH70_02080 and AH70_02070 on contig 14, respectively) but not in the *P. pentosaceus* ATCC 25745 genome sequence. AH70_09615 and AH70_02080 were annotated as an acyl-phosphate glycerol 3-phospate and a glycosyl transferase,

respectively. Both PECL_1950 and PECL_1952 are members of the lysophospholipid acyltransferase superfamily, which contains acyltransferases of *de novo* and remodeling pathways of glycerophospholipid biosynthesis. Finally, AH70_02070 contained a family 8 glycosyl transferase domain that catalyzes the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds.

pPECL-8 (*P. claussenii* ATCC BAA-344^T)



Contigs 70, 14, and 97 of P. damnosus LMG 28219

Figure 19: Comparison of plasmid pPECL-8 from P. claussenii *ATCC BAA 344^T and contigs 70, 14, and 90 of* P. damnosus *LMG 28219*

The Artemis comparison tool (Carver *et al.*, 2005) alignment was based on a BLASTn comparison using default settings. The red lines indicate that the aligned regions had the same orientation, whereas the blue lines indicate that the aligned regions were inversely oriented. The *horA* gene and its surrounding genes are indicated as grey blocks [from left to right: AH70_091615 (glycosyl transferase), AH70_09610 (glycosyl transferase), AH70_02080 (acyl-phoshate glycerol 3-phospate), HorA: AH70_02075 (multidrug transporter), and AH70_02070 (glycosyl transferase)]. p: plasmid.

Next to the *horA* gene, the *hitA* gene encodes a divalent cation transporter and confers hop resistance by importing manganese and thereby counteracting proton gradient dissipation (Hayashi *et al.*, 2001). Orthologs of the *hitA* gene of *Lb. brevis* L5784 (AB035808) were found in the genomes of *P. damnosus* LMG 28219 (AH70_05540), *P. claussenii* ATCC BAA 344^T (AEV94625.1), and *P. pentosaceus* ATCC 25745 (YP_805164). The reason for the presence of the *hitA* gene in *P. pentosaceus* ATCC 25745 of plant origin remains unclear. Furthermore, HorC has been suggested to be a proton motive force-dependent multidrug transporter, whose expression is under control of the HorB transcriptional regulator (Iijima *et al.*, 2006). Orthologs of the *horC* gene of *Lactobacillus backii* LMG 23555 (BAF56899.1) were found on contig 18 (AH70_03090) in the *P. damnosus* LMG 28219 and *P. claussenii* ATCC BAA 344^T genome sequences (AEV95756.1), but not in that of *P. pentosaceus* ATCC 25745. Finally, *horB*, *bsrA*, and *bsrB* homologs were not found in the draft genome sequence of *P. damnosus* LMG 28219.

Folate Biosynthesis

The beer isolates *P. damnosus* LMG 28219 and *P. claussenii* ATCC BAA-344^T contained a set of genes involved in folate biosynthesis that were absent in the genome of *P. pentosaceus* ATCC 25745 (of plant origin). These included a 5,10-methylenetetrahydrofolate reductase (EC 1.5.1.20, AH70_04245 on contig 29), a dihydropteroate synthase (EC 2.5.1.15, AH70_05560 on contig 4), a 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3, AH70_-05580 on contig 4), a GTP cyclohydrolase I type 1 (EC 3.5.4.16, AH70_05575 on contig 4), and a dihydroneopterin aldolase (EC 4.1.2.25, AH70_05585 on contig 4).

Because folate-dependent formylation of the initiator tRNA is a hallmark of bacterial translation and because bacteria cannot import formylmethionyl-tRNA, folate is essential for bacterial growth (de Crecy-Lagard *et al.*, 2007). Most bacteria make folate *de novo*, starting from GTP and chorismate (de Crecy-Lagard *et al.*, 2007). The first enzyme in *de novo* folate biosynthesis is GTP cyclohydrolase I that catalyzes a complex reaction, in which the five-membered imidazole ring of GTP is opened and a six-membered dihydropyrazine ring is formed (Figure 20). The resulting 7,8-dihydroneopterin triphosphate is then converted into the corresponding monophosphate by a specific pyrophosphatase. A pyrophosphatase was found in the genome of *P. claussenii* ATCC BAA-344^T but was absent in the draft genome sequence of *P. damnosus* LMG 28219. BLASTp of the pyrophosphatase gene against the predicted CDS amino acid sequences of *P. damnosus* LMG 28219 did

not reveal the presence of a homologous gene. Possibly, the gene performing this function in *P. damnosus* LMG 28219 has not been previously identified or is missing due to non-orthologous gene replacement. Alternatively, some missing information in the draft genome sequence of *P. damnosus* LMG 28219 may account for this gap in the folate biosynthesis pathway. Dihydroneopterin aldolase subsequently releases glycoaldehyde to produce 6-hydroxymethyl-7,8-dihydropterin, which is then pyrophosphorylated by hydroxymethyldihydropterin pyrophosphokinase. 6-Hydroxymethyl-7,8-dihydropterin pyrophosphate and 4-aminobenzoic acid (pABA) moieties are condensed by dihydropteroate synthase and this results in the production of dihydropteroate. The enzymes involved in the conversion of dihydropteroate into tetrahydrofolate-polyglutamate are shared by the three genomes analyzed. These differences in folate biosynthesis potential between the *P. damnosus* LMG 28219, *P. claussenii* ATCC BAA-344^T, and *P. pentosaceus* ATCC 25745 genomes may be explained by the environmental acquisition of folate by the latter strain. Plants produce folate *de novo*, so possibly *P. pentosaceus* ATCC 25745 harbors folate transporters to import folate produced by its host and lost genes involved in folate biosynthesis during evolution.

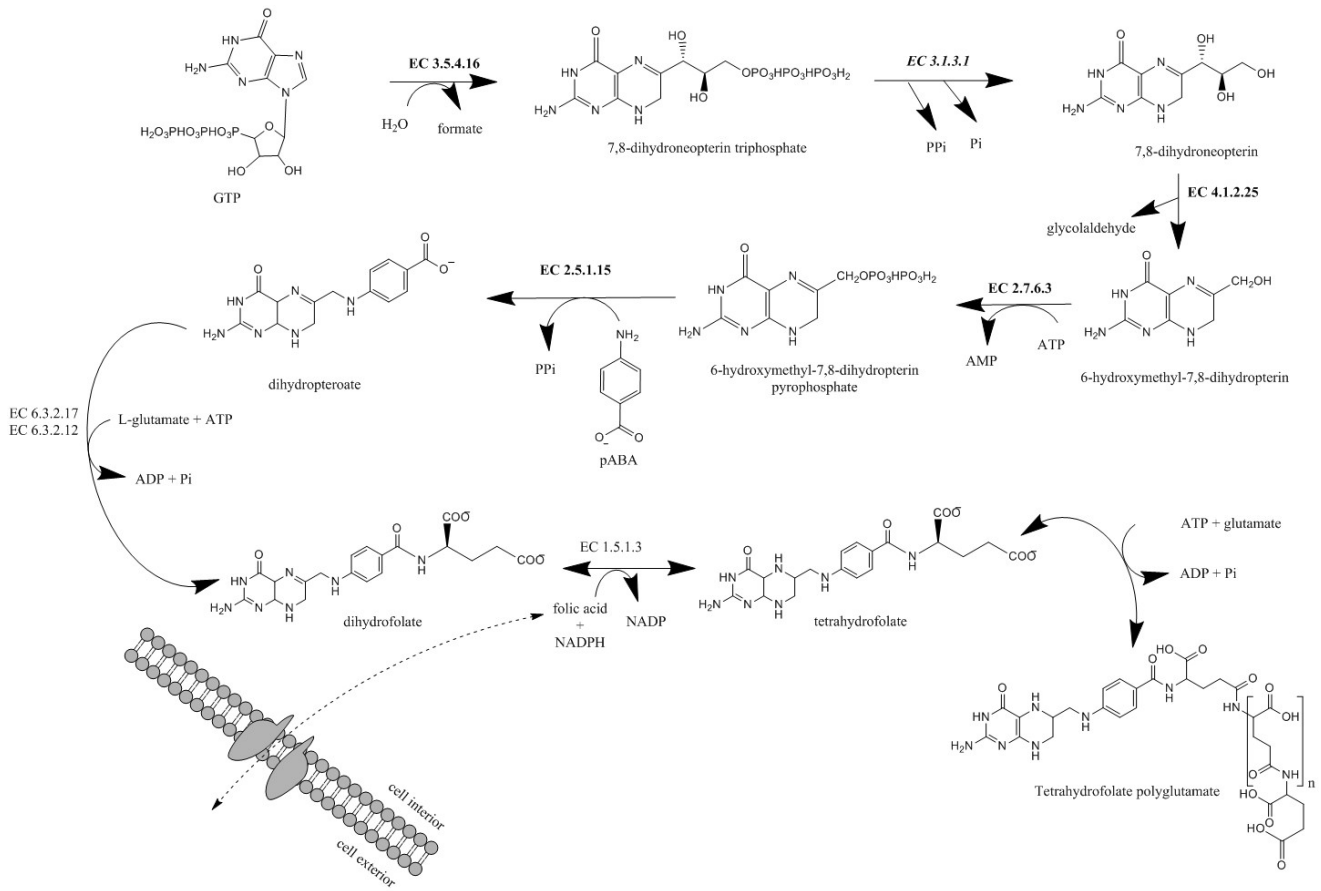


Figure 20: Folate biosynthesis pathway reconstruction

The enzymes found in both *P. damnosus* LMG 28219 and *P. claussenii* ATCC BAA 344^T are highlighted in bold, with the exception of EC 3.1.3.1 that was only found in the genome of *P. claussenii* ATCC BAA 344^T (highlighted in bold and italic). The remaining enzymes were shared by *P. damnosus* LMG 28219, *P. claussenii* ATCC BAA 344^T, and *P. pentosaceus* ATCC 25745. The mechanism of the reaction highlighted with a dashed arrow is uncertain and needs further investigation. The figure was constructed using the ChemBioDraw software v. 13.0 (Perkin Elmer Inc., Waltham, MA, USA). ATP: adenosine triphosphate, ADP: adenosine diphosphate, AMP: adenosine monophosphate, NADP(H): nicotinamide adenine dinucleotide phosphate, P(P)i: (pyro)phosphate, EC: enzyme commission, GTP: guanoside triphospate.

EPS Production

Laboratory experiments indicated that *P. damnosus* LMG 28219 produces EPS. The ability of a strain to produce EPS is not directly correlated to its ability to reside in beer but probably has importance in biofilm formation (Pittet *et al.*, 2011), thereby enabling persistence in the brewery environment. The EPS produced by *P. claussenii* ATCC BAA 344^T is a high-molecular-mass β-glucan produced by the action of a transmembrane glycosyl transferase (*gtf*) gene. This gene is located on plasmid pPECL-7, which is not essential for growth in beer (Pittet *et al.*, 2013). The fibrillar polymer consists of a trisaccharide repeating unit with a β-1,3-linked glucose backbone and branches made up of single β-1,2-linked D-glucopyranosyl residues. Walling and colleagues (2005) reported a glucosyl transferase gene (*dps*) in *P. damnosus* IOEB8801, originating from wine, that produces a linear backbone of 3-β-D-glucose-1 moieties. Surprisingly, no homologies towards the *gtf* and *dps* genes were found in the draft genome sequence of *P. damnosus* LMG 28219. Yet, contig 56 harbored a CDS (AH70_07835), containing a glycosyl transferase (group 2) domain, which may be involved in EPS production. OrthoMCL analysis indicated that this CDS was unique to *P. damnosus* LMG 28219. Other CDS candidates involved in EPS production by *P. damnosus* LMG 28219 were AH70_01405, AH70_01410, and AH70_01415, which were all predicted to be present on contig 122. These proteins did not show amino acid sequence homologies of more than 70% towards the proteomes of *P. claussenii* ATCC BAA 344^T and *P. pentosaceus* ATCC 25745. AH70_01405 is an EPS biosynthesis protein, consisting of an AAA domain containing a P-loop, and showed similarities towards the capsular EPS family protein (EHO53752) found in the genome of *Lactobacillus kisonensis* F0435, originating from the human oral cavity (Chen *et al.*, 2010). AH70_01410 is a lipopolysaccharide biosynthesis domain that is involved in the biosynthesis of EPS. This protein showed more than 70% protein sequence homology towards the chain length determinant protein (EEI18212) of *Lb. buchneri* ATCC 11577, isolated from the human oral cavity (Tilden & Svec, 1952). Finally, AH70_01415 harbored a cell envelope-related transcriptional attenuator domain, which describes a domain of unknown function that is found in the predicted extracellular domain of a number of putative membrane-bound proteins (Zdobnov & Apweiler, 2001). One of those is CpsA, a putative regulatory protein involved in EPS biosynthesis (Zdobnov & Apweiler, 2001). AH70_01415 showed more than 70% protein sequence homology towards a cell envelope-like function transcriptional attenuator common domain protein (EEI18211) of *Lb. buchneri* ATCC 11577 and a biofilm regulatory protein of *Lb. kisonensis* F0435 (see above). Besides the presence of proteins potentially involved in EPS production in *P. damnosus* LMG 28219 (as discussed above), the mechanism of EPS production in *P. damnosus* LMG 28219 remains unclear.

Genes Involved in Oxidative Stress Response

Pittet and coworkers (2013) found a set of highly transcribed genes in *P. claussenii* ATCC BAA 344^T during growth in beer as a response to the oxidative stress imposed by hops. These genes are manganese transport proteins, methionine sulfoxide reductases MrsA and MsrB as well as other metal transport and homeostasis proteins. Orthologs of these genes were found in the genomes of *P. damnosus* LMG 28219 and *P. pentosaceus* ATCC 25745, indicating that the presence of these genes is not unique to strains capable of growing in beer. These results indicate that growth in beer is a multifactorial reaction of the cell towards a challenging environment, not only involving the presence of specific genes but also the up-and down-regulation of specific sets of genes.

Functions Enriched in the Genomes of Bacteria Originating from Beer

Next to functions unique to bacteria originating from beer, enriched functions could also provide insight into the mechanisms of niche adaptation. The four most abundant non-core LaCOGs (COG 2814, COG1309, COG0789, and COG2826, discussed above) in the *P. damnosus* LMG 28219 draft genome were also present in the *P. claussenii* ATCC BAA-344^T and *P. pentosaceus* ATCC 25745 genomes, but were enriched only in the genomes of the bacterial strains originating from beer (*i.e.*, *P. damnosus* LMG 28219 and *P. claussenii* ATCC BAA-344^T) (Figure 21). Other enriched functions were COG1846 (transcriptional regulators), COG0596 (predicted hydrolases or acyltransferases of the alpha/beta hydrolase superfamily), and COG0745 (response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain). The enrichment of the above-mentioned functions in bacteria growing in beer was further substantiated when analyzing additional genome sequences of the genus *Pediococcus* and *Lactobacillus*. Many COGs that were enriched in bacteria originating from beer are involved in transcriptional regulation, indicating that genes can be up- or down-regulated in hop-stressed cells. Currently, it is not clear in which processes these regulators are involved but this may be revealed by transcriptome analyses.

		No beer origin							Beer origin			
COG	**Name**	*L. delbrueckii bulgaricus* ATCC 11842	*P. acidilactici* 7_4	*P. acidilactici* DSM 20284	*P. acidilactici* MA 18/5M	*P. acidilactici* NGRI 0510Q	*P. pentosaceus* ATCC 25745	*P. pentosaceus* IE-3	*P. claussenii* ATCC BAA-344^T	*P. damnosus* LMG 28219	*L. malefermentans* KCTC 3548	*L. rhamnosus* ATCC 8530
COG2814	Arabinose efflux permease	8	18	19	18	19	16	16	22	24	23	26
COG1309	Transcriptional regulator	2	4	4	4	7	5	7	23	14	13	16
COG0789	Predicted transcriptional regulator	2	6	5	6	5	5	4	8	12	9	6
COG2826	Transposase and inactivated derivatives, IS30 family	4	0	1	1	1	1	2	4	12	10	4
COG1846	Transcriptional regulator	1	5	5	5	6	6	7	10	11	16	9
COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	3	2	2	2	3	2	2	4	8	4	8
COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	5	5	5	5	5	5	5	7	8	6	12

Figure 21: Overview of the functions enriched in the genomes of LAB of beer origin compared to those of LAB not originating from beer

The number of CDSs in the different COGs are listed. The color gradient ranges from white to black with increasing number of CDS and applies per row. IMG-ER was used for the determination of the number of CDSs. COG: cluster of orthologous groups, *Lb. Lactobacillus*, IS: insertion sequence.

5.2.4 Conclusions

The draft genome of *P. damnosus* LMG 28219 and its comparative analysis with other pediococci provided insights into the adaptation of this strain to the beer environment. These adaptations included the presence of the *horA* gene and its surrounding genes, genes involved in *de novo* folate biosynthesis, genes involved in the production of EPS, and the enrichment of functions related to transcriptional regulation. Furthermore, the results presented in this study will enable future transcriptome analysis of *P. damnosus* LMG 28219, which can provide additional insights into its metabolic activities.

5.2.5 Acknowledgments

The authors acknowledge the financial support of FWO-Flanders, BOF project and the Vrije Universiteit Brussel (SRP, IRP, and IOF projects), and of the Hercules Foundation. In addition, the authors would like to acknowledge the brewery involved and Charlotte Peeters and Bart Verheyde for the bioinformatics support.

5.2.6 Supplementary Material

The raw sequence data received from BaseClear BV were deposited at the SRA of GenBank (accession number SRP035530). This Whole-genome shotgun project was deposited at DDBJ/EMBL/GenBank under the accession number JANK00000000 after automatic annotation by PGAAP and manual curation of the genes of interest.

Supplementary Material 5.8: Overview of the G+C content and the length of the contigs of the draft genome of P. damnosus *LMG 28219*

contig no.	GC-content (%)	length (bp)	contig no.	GC-content (%)	length (bp)	contig no.	GC-content (%)	length (bp)
1	41.83	502	62	39.12	25631	123	38.10	1756
2	35.96	29429	63	38.10	1614	124	37.86	19366
3	35.66	3738	64	36.64	52573	125	42.39	2958
4	39.18	17052	65	38.67	22379	126	37.48	1174
5	34.15	9243	66	37.62	30560	127	34.09	20107
6	37.64	16203	67	38.29	1319	128	39.60	9231
7	39.31	53329	68	38.15	53951	129	37.49	5865
8	36.84	11309	69	36.79	15596	130	38.24	2691
9	41.49	1569	70	44.45	1982	131	39.49	5242
10	39.66	24545	71	37.03	18224	132	38.02	2967
11	37.39	27167	72	36.32	18506	133	43.34	953
12	37.00	5940	73	38.73	7361	134	34.51	6286
13	44.28	1154	74	38.62	35374	135	34.29	630
14	41.30	6274	75	45.65	5490	136	37.48	22993
15	43.70	1906	76	38.60	10142	137	36.68	3539
16	49.05	1790	77	39.44	21192	138	38.83	703
17	36.98	15880	78	44.06	6518	139	39.28	6973
18	42.10	14520	79	42.35	6140	140	39.10	4059
19	37.06	39520	80	37.70	29394	141	32.78	6114
20	37.61	15298	81	39.17	43485	142	34.46	711
21	38.42	10835	82	38.44	13240	143	34.61	1598
22	39.32	5570	83	38.08	17031	144	37.44	2094
23	37.66	28945	84	39.63	10334	145	41.14	6624
24	35.63	10409	85	37.86	14162	146	34.62	11446
25	37.24	1450	86	41.89	2134	147	38.32	6765
26	37.53	22807	87	36.99	2933	148	38.77	1251
27	39.68	41367	88	41.39	720	149	33.61	2565
28	38.52	3580	89	33.63	5376	150	39.22	2675
29	37.66	112846	90	38.38	7835	151	40.74	7364
30	37.97	23931	91	40.83	5787	152	39.43	8752
31	36.80	8354	92	43.40	735	153	36.07	1594
32	36.98	67552	93	43.89	565	154	34.98	1921
33	39.12	1337	94	46.11	1028	155	35.21	676
34	37.68	6696	95	37.95	1149	156	34.56	2584
35	39.11	42230	96	35.55	7300	157	41.35	1468
36	47.74	1906	97	41.72	20292	158	41.37	556
37	44.61	1242	98	39.53	9354	159	30.99	755
38	41.52	9357	99	37.68	9719	160	37.77	5809
39	38.53	1547	100	34.96	1533	161	46.71	1824
40	37.76	1213	101	37.85	21944	162	39.71	1244
41	36.10	2906	102	40.15	24659	163	39.02	3350
42	39.01	905	103	37.56	13572	164	35.96	2912
43	36.94	25254	104	37.97	31232	165	38.27	2710
44	46.78	6624	105	38.67	2801	166	35.70	2678
45	37.45	23239	106	40.88	2275	167	36.36	1276
46	39.37	68863	107	36.22	7228	168	39.08	16295
47	40.25	20521	108	44.34	857	169	38.69	4125
48	36.41	11759	109	41.12	642	170	41.54	5611
49	37.67	36252	110	39.39	12345	171	32.33	532
50	37.34	31306	111	40.97	3937	172	30.56	1705
51	37.11	16670	112	36.98	933	173	39.62	785
52	39.53	55967	113	41.76	771	174	36.17	1247
53	38.50	59279	114	38.19	40025	175	36.95	609
54	38.44	23452	115	39.38	1097	176	49.68	1872
55	36.90	18764	116	39.03	16480	177	43.89	1399
56	38.94	35768	117	37.61	1550	178	38.88	1003
57	48.97	3286	118	39.27	4077	179	45.93	553
58	36.80	22440	119	38.40	22575	180	33.47	983
59	36.99	22335	120	37.74	12548	181	41.15	593
60	40.23	17760	121	38.45	3555	182	32.04	774
61	37.60	8570	122	32.87	10225	183	34.59	873

Supplementary Material 5.9: Shared and unique orthologous proteins of P. damnosus *LMG 28219,* P. claussenii *ATCC BAA 344^T, and* P. pentosaceus *ATCC 25745*

clau: *P. claussenii* ATCC BAA 344^T, ldam: *P. damnosus* LMG 28219, pent: *P. pentosaceus* ATCC 25745

is0000: clau|AEV94415.1 clau|AEV95932.1 clau|AEV95653.1 ldam|AH70_02430 ldam|AH70_03715 ldam|AH70_03720 ldam|AH70_03725 ldam|AH70_07075 ldam|AH70_-01065 pent|YP_805047 pent|YP_803934
is0001: clau|AEV95460.1 clau|AEV95547.1 clau|AEV95762.1 ldam|AH70_05025 ldam|AH70_06205 ldam|AH70_09260 pent|YP_804136 pent|YP_804474 pent|YP_-804906 pent|YP_804593
is0002: clau|AEV96110.1 clau|AEV96124.1 clau|AEV96156.1 clau|AFD55023.1 ldam|AH70_03040 ldam|AH70_05620 ldam|AH70_11090 ldam|AH70_11270
is0003: clau|AEV94465.1 clau|AEV96016.1 ldam|AH70_01170 ldam|AH70_04480 ldam|AH70_08070 pent|YP_803658 pent|YP_803822
is0004: clau|AEV96148.1 clau|AEV96190.2 ldam|AH70_00605 ldam|AH70_11200 ldam|AH70_03075 ldam|AH70_03865 pent|YP_805170
is0005: ldam|AH70_01040 ldam|AH70_02645 ldam|AH70_10065 ldam|AH70_02650 ldam|AH70_07425 ldam|AH70_08475 ldam|AH70_09300
is0006: clau|AEV94344.1 clau|AEV95200.1 ldam|AH70_04185 ldam|AH70_05235 pent|YP_803558 pent|YP_804457
is0007: clau|AEV94407.1 clau|AEV95586.1 pent|YP_803794 pent|YP_803942 ldam|AH70_00840 ldam|AH70_08055
is0008: clau|AEV94515.1 clau|AEV95780.1 clau|AEV96066.1 ldam|AH70_01320 pent|YP_803931 ldam|AH70_06130
is0009: clau|AEV95463.1 clau|AEV95699.1 ldam|AH70_00670 ldam|AH70_09265 pent|YP_803905 pent|YP_804133
is0010: ldam|AH70_01495 ldam|AH70_08760 ldam|AH70_06900 ldam|AH70_05735 ldam|AH70_02020 ldam|AH70_10145
is0011: clau|AEV94632.1 ldam|AH70_01940 ldam|AH70_10475 pent|YP_804995 clau|AEV95979.1 pent|YP_805114
is0012: clau|AEV94454.1 clau|AEV96074.1 clau|AEV95794.1 pent|YP_804423 ldam|AH70_06030
is0013: clau|AEV94821.1 clau|AEV94913.1 pent|YP_803615 pent|YP_803966 ldam|AH70_09875
is0014: clau|AEV96109.1 clau|AEV96157.1 clau|AEV96123.1 clau|AFD55024.1 ldam|AH70_03620
is0015: ldam|AH70_02400 ldam|AH70_04355 clau|AEV96063.1 pent|YP_804927 pent|YP_805296
is0016: ldam|AH70_03055 ldam|AH70_07765 ldam|AH70_07795 clau|AEV94710.1 pent|YP_804054
is0017: ldam|AH70_08370 ldam|AH70_09835 pent|YP_803719 pent|YP_805210 clau|AEV96005.1
is0018: pent|YP_803703 pent|YP_804555 pent|YP_804045 pent|YP_804079 pent|YP_805062
is0019: clau|AEV94860.1 ldam|AH70_07675 clau|AEV95588.1 ldam|AH70_01525 pent|YP_803940
is0020: clau|AEV94602.1 clau|AEV96113.1 ldam|AH70_01240 pent|YP_804037
is0021: clau|AEV94627.1 clau|AEV94669.1 ldam|AH70_04530 pent|YP_803636
is0022: clau|AEV94628.1 clau|AEV95664.1 ldam|AH70_04535 pent|YP_803637
is0023: clau|AEV94795.1 clau|AFD55028.1 ldam|AH70_01045 ldam|AH70_03875
is0024: clau|AEV95797.1 clau|AEV96182.1 ldam|AH70_08980 pent|YP_805108
is0025: clau|AEV95888.1 clau|AEV95889.1 pent|YP_805266 ldam|AH70_04700
is0026: clau|AEV96083.1 clau|AEV96093.1 clau|AEV96188.1 ldam|AH70_10315
is0027: clau|AEV96119.1 clau|AEV96162.1 ldam|AH70_03600 ldam|AH70_00575
is0028: ldam|AH70_00820 ldam|AH70_01285 clau|AEV96115.1 pent|YP_804032
is0029: ldam|AH70_01030 ldam|AH70_02720 clau|AEV94534.1 pent|YP_805122
is0030: ldam|AH70_01035 ldam|AH70_02075 clau|AEV96194.1 pent|YP_805121
is0031: ldam|AH70_08835 ldam|AH70_10820 ldam|AH70_03360 ldam|AH70_06115
is0032: ldam|AH70_01270 ldam|AH70_02155 ldam|AH70_03595 clau|AEV96118.1
is0033: ldam|AH70_01335 ldam|AH70_06980 clau|AEV95990.1 pent|YP_805277
is0034: ldam|AH70_01420 ldam|AH70_01835 ldam|AH70_03395 clau|AEV96185.1
is0035: ldam|AH70_01660 ldam|AH70_01665 clau|AEV95481.1 pent|YP_804113
is0036: ldam|AH70_01810 ldam|AH70_03840 clau|AEV95485.1 pent|YP_804104
is0037: ldam|AH70_01850 ldam|AH70_03095 clau|AEV96173.1 pent|YP_804033
is0038: ldam|AH70_02050 ldam|AH70_06475 ldam|AH70_11060 pent|YP_803954
is0039: ldam|AH70_02480 ldam|AH70_08740 clau|AEV96170.1 pent|YP_805144
is0040: ldam|AH70_02515 ldam|AH70_10990 ldam|AH70_01120 ldam|AH70_05655
is0041: ldam|AH70_02670 ldam|AH70_10425 clau|AEV95240.1 pent|YP_804410
is0042: ldam|AH70_04015 ldam|AH70_06460 clau|AEV95437.1 pent|YP_804167
is0043: ldam|AH70_05435 ldam|AH70_07100 clau|AEV94712.1 pent|YP_803867
is0044: ldam|AH70_06240 ldam|AH70_11190 clau|AEV96052.1 pent|YP_804728
is0045: ldam|AH70_06530 ldam|AH70_10445 clau|AEV94387.1 pent|YP_803603
is0046: ldam|AH70_07095 ldam|AH70_08780 ldam|AH70_10125 clau|AEV96025.1
is0047: ldam|AH70_07210 ldam|AH70_10440 ldam|AH70_07245 clau|AEV95904.1
is0048: ldam|AH70_07235 ldam|AH70_09690 clau|AEV94930.1 pent|YP_805282
is0049: ldam|AH70_07775 ldam|AH70_10230 clau|AEV95913.1 pent|YP_804465
is0050: ldam|AH70_08310 ldam|AH70_11215 clau|AEV96199.1 pent|YP_804726
is0051: ldam|AH70_09460 ldam|AH70_10290 clau|AEV94894.1 pent|YP_804504
is0052: ldam|AH70_10005 ldam|AH70_10995 clau|AEV95471.1 pent|YP_804126
is0053: pent|YP_803729 pent|YP_805064 clau|AEV96015.1 ldam|AH70_04485
is0054: pent|YP_803842 pent|YP_803919 clau|AEV94745.1 ldam|AH70_07525
is0055: pent|YP_803850 pent|YP_805265 clau|AEV94498.1 ldam|AH70_07445
is0056: pent|YP_803953 pent|YP_805169 clau|AEV94766.1 ldam|AH70_07355
is0057: pent|YP_804144 pent|YP_804735 clau|AEV95025.1 ldam|AH70_05545
is0058: pent|YP_804251 pent|YP_804532 clau|AEV95150.1 ldam|AH70_09565
is0059: pent|YP_804280 pent|YP_804506 clau|AEV94889.1 ldam|AH70_09485
is0060: pent|YP_804402 pent|YP_805227 clau|AEV94555.1 ldam|AH70_07970
is0061: ldam|AH70_06575 ldam|AH70_10765 ldam|AH70_06865 ldam|AH70_06580
is0062: clau|AEV94358.1 ldam|AH70_10660 pent|YP_803588 pent|YP_805105
is0063: clau|AEV94359.1 ldam|AH70_05550 pent|YP_805104 pent|YP_803824
is0064: clau|AEV94383.1 pent|YP_803653 clau|AEV96058.1 ldam|AH70_06960
is0065: clau|AEV94389.1 pent|YP_804544 clau|AEV95765.1 ldam|AH70_09650

is0066: clau|AEV94392.1 ldam|AH70_09580 clau|AEV96114.1 pent|YP_804038
is0067: clau|AEV94438.1 pent|YP_804361 clau|AEV94674.1 ldam|AH70_10550
is0068: clau|AEV94440.1 pent|YP_804363 clau|AEV94675.1 ldam|AH70_10555
is0069: clau|AEV94441.1 pent|YP_804364 clau|AEV94676.1 ldam|AH70_10560
is0070: clau|AEV94442.1 pent|YP_804365 clau|AEV94677.1 ldam|AH70_10565
is0071: clau|AEV94484.1 ldam|AH70_04720 pent|YP_803828 pent|YP_803639
is0072: clau|AEV94512.1 ldam|AH70_06815 ldam|AH70_06955 pent|YP_805058
is0073: clau|AEV94561.1 ldam|AH70_07395 clau|AEV95958.1 pent|YP_803596
is0074: clau|AEV94568.1 ldam|AH70_03440 clau|AEV95456.1 pent|YP_804140
is0075: clau|AEV94647.1 ldam|AH70_06860 pent|YP_803577 pent|YP_804981
is0076: clau|AEV94708.1 ldam|AH70_04665 pent|YP_805110 pent|YP_804623
is0077: clau|AEV94755.1 ldam|AH70_01855 clau|AEV95802.1 pent|YP_804106
is0078: clau|AEV94807.1 ldam|AH70_04870 pent|YP_804948 pent|YP_803750
is0079: clau|AEV94876.1 ldam|AH70_03315 clau|AEV96022.1 pent|YP_805244
is0080: clau|AEV95036.1 pent|YP_804722 clau|AEV95668.1 ldam|AH70_09680
is0081: clau|AEV95338.1 ldam|AH70_11130 pent|YP_804312 ldam|AH70_11135
is0082: clau|AEV95468.1 ldam|AH70_11015 pent|YP_803674 pent|YP_805067
is0083: clau|AEV95486.1 ldam|AH70_05300 pent|YP_805066 pent|YP_805071
is0084: clau|AEV95798.1 ldam|AH70_08975 clau|AEV96183.1 pent|YP_805109
is0085: clau|AEV95831.1 ldam|AH70_01710 pent|YP_803601 pent|YP_805148
is0086: clau|AEV95835.1 ldam|AH70_00940 clau|AEV95855.1 pent|YP_805175
is0087: clau|AEV95860.1 ldam|AH70_02675 clau|AEV96096.1 pent|YP_803702
is0088: clau|AEV95884.1 ldam|AH70_03500 pent|YP_804540 pent|YP_804051
is0089: clau|AEV96192.1 ldam|AH70_09610 pent|YP_803759 ldam|AH70_08745
is0090: clau|AEV94734.1 clau|AEV94735.1 pent|YP_805068
is0091: clau|AEV95602.1 clau|AEV95610.1 clau|AEV95603.1
is0092: clau|AEV95615.1 clau|AEV95621.1 clau|AEV95617.1
is0093: clau|AEV95754.1 clau|AEV95941.1 ldam|AH70_03640
is0094: clau|AEV96092.1 clau|AEV96189.1 ldam|AH70_05630
is0095: clau|AEV96099.1 clau|AEV96179.1 ldam|AH70_03145
is0096: clau|AEV96105.1 clau|AEV96112.1 ldam|AH70_08970
is0097: clau|AEV96121.1 clau|AFD55026.1 ldam|AH70_01860
is0098: clau|AEV96140.1 clau|AFD55022.1 ldam|AH70_01815
is0099: clau|AEV96142.1 clau|AEV96206.1 ldam|AH70_11255
is0100: ldam|AH70_00345 ldam|AH70_11230 clau|AEV96203.1
is0101: ldam|AH70_00835 ldam|AH70_03900 ldam|AH70_03855
is0102: ldam|AH70_00865 ldam|AH70_10190 pent|YP_804449
is0103: ldam|AH70_00985 ldam|AH70_01290 ldam|AH70_07910
is0104: ldam|AH70_01260 ldam|AH70_01515 clau|AEV96031.1
is0105: ldam|AH70_01305 ldam|AH70_02490 ldam|AH70_01745
is0106: ldam|AH70_01920 ldam|AH70_05480 ldam|AH70_02555
is0107: ldam|AH70_02340 ldam|AH70_08735 ldam|AH70_02485
is0108: ldam|AH70_03165 ldam|AH70_04585 ldam|AH70_04610
is0109: ldam|AH70_03265 ldam|AH70_03310 clau|AEV94877.1
is0110: ldam|AH70_04255 ldam|AH70_07600 clau|AEV95845.1
is0111: ldam|AH70_04670 ldam|AH70_04920 clau|AEV94852.1
is0112: ldam|AH70_05350 ldam|AH70_09665 pent|YP_803604
is0113: ldam|AH70_05470 ldam|AH70_06015 ldam|AH70_05615
is0114: ldam|AH70_05725 ldam|AH70_08765 ldam|AH70_07160
is0115: ldam|AH70_07085 ldam|AH70_08755 clau|AEV95658.1
is0116: ldam|AH70_07270 ldam|AH70_07840 ldam|AH70_07845
is0117: ldam|AH70_08250 ldam|AH70_10705 pent|YP_804836
is0118: pent|YP_803711 pent|YP_804553 ldam|AH70_01325
is0119: pent|YP_804248 pent|YP_804536 ldam|AH70_02630
is0120: pent|YP_804259 pent|YP_804521 ldam|AH70_09535
is0121: clau|AEV94340.1 ldam|AH70_04205 pent|YP_803554
is0122: clau|AEV94341.1 ldam|AH70_04200 pent|YP_803555
is0123: clau|AEV94342.1 ldam|AH70_04195 pent|YP_803556
is0124: clau|AEV94343.1 ldam|AH70_04190 pent|YP_803557
is0125: clau|AEV94345.1 ldam|AH70_04180 pent|YP_803559
is0126: clau|AEV94346.1 ldam|AH70_10645 pent|YP_803560
is0127: clau|AEV94347.1 ldam|AH70_10650 pent|YP_803561
is0128: clau|AEV94348.1 ldam|AH70_10655 pent|YP_803562
is0129: clau|AEV94349.1 ldam|AH70_10675 pent|YP_803563
is0130: clau|AEV94350.1 ldam|AH70_10680 pent|YP_803564
is0131: clau|AEV94351.1 ldam|AH70_10685 pent|YP_803565
is0132: clau|AEV94352.1 ldam|AH70_10690 pent|YP_803566
is0133: clau|AEV94354.1 ldam|AH70_01115 pent|YP_803567
is0134: clau|AEV94357.1 ldam|AH70_01220 pent|YP_803569
is0135: clau|AEV94361.1 ldam|AH70_01695 pent|YP_803826
is0136: clau|AEV94368.1 ldam|AH70_09525 pent|YP_804859
is0137: clau|AEV94370.1 ldam|AH70_03755 pent|YP_805224
is0138: clau|AEV94371.1 ldam|AH70_03760 pent|YP_805223
is0139: clau|AEV94372.1 ldam|AH70_03765 pent|YP_805222
is0140: clau|AEV94373.1 ldam|AH70_03770 pent|YP_805221
is0141: clau|AEV94374.1 ldam|AH70_03775 pent|YP_805220
is0142: clau|AEV94375.1 ldam|AH70_03785 pent|YP_805218
is0143: clau|AEV94378.1 ldam|AH70_01455 pent|YP_803692
is0144: clau|AEV94379.1 ldam|AH70_04440 pent|YP_803651
is0145: clau|AEV94380.1 ldam|AH70_04445 pent|YP_803652
is0146: clau|AEV94382.1 ldam|AH70_00855 pent|YP_804860
is0147: clau|AEV94384.1 ldam|AH70_04460 pent|YP_803654
is0148: clau|AEV94385.1 ldam|AH70_04470 pent|YP_805171
is0149: clau|AEV94386.1 ldam|AH70_04390 pent|YP_803776
is0150: clau|AEV94390.1 ldam|AH70_10715 pent|YP_804116
is0151: clau|AEV94391.1 ldam|AH70_10795 pent|YP_803778
is0152: clau|AEV94393.1 ldam|AH70_11120 pent|YP_805188
is0153: clau|AEV94394.1 ldam|AH70_11280 pent|YP_803779
is0154: clau|AEV94395.1 ldam|AH70_11285 pent|YP_803780
is0155: clau|AEV94399.1 ldam|AH70_11290 pent|YP_803781
is0156: clau|AEV94400.1 ldam|AH70_11295 pent|YP_803782
is0157: clau|AEV94401.1 ldam|AH70_11300 pent|YP_803783
is0158: clau|AEV94402.1 ldam|AH70_11305 pent|YP_803784
is0159: clau|AEV94408.1 ldam|AH70_02170 pent|YP_803795
is0160: clau|AEV94409.1 ldam|AH70_07265 pent|YP_804421
is0161: clau|AEV94412.1 ldam|AH70_02180 pent|YP_803796
is0162: clau|AEV94413.1 ldam|AH70_02185 pent|YP_803797
is0163: clau|AEV94414.1 ldam|AH70_02190 pent|YP_803798
is0164: clau|AEV94416.1 ldam|AH70_10070 pent|YP_803801
is0165: clau|AEV94418.1 ldam|AH70_10095 pent|YP_803804
is0166: clau|AEV94419.1 ldam|AH70_10100 pent|YP_803805
is0167: clau|AEV94420.1 ldam|AH70_10110 pent|YP_803806
is0168: clau|AEV94421.1 ldam|AH70_10115 pent|YP_803807
is0169: clau|AEV94423.1 ldam|AH70_10155 pent|YP_803809
is0170: clau|AEV94424.1 ldam|AH70_10160 pent|YP_803810
is0171: clau|AEV94425.1 ldam|AH70_10180 pent|YP_803811
is0172: clau|AEV94426.1 ldam|AH70_01255 pent|YP_803812
is0173: clau|AEV94427.1 ldam|AH70_01245 pent|YP_803813
is0174: clau|AEV94435.1 ldam|AH70_03470 pent|YP_804358
is0175: clau|AEV94448.1 ldam|AH70_00910 pent|YP_803609
is0176: clau|AEV94450.1 ldam|AH70_04620 pent|YP_805208
is0177: clau|AEV94451.1 ldam|AH70_01235 pent|YP_803816
is0178: clau|AEV94452.1 ldam|AH70_01230 pent|YP_803817
is0179: clau|AEV94453.1 ldam|AH70_01225 pent|YP_803818
is0180: clau|AEV94455.1 ldam|AH70_04110 pent|YP_804838
is0181: clau|AEV94462.1 ldam|AH70_01190 pent|YP_803819
is0182: clau|AEV94463.1 ldam|AH70_01180 pent|YP_803820
is0183: clau|AEV94464.1 ldam|AH70_01175 pent|YP_803821
is0184: clau|AEV94466.1 ldam|AH70_01165 pent|YP_803823
is0185: clau|AEV94469.1 ldam|AH70_06480 pent|YP_803679
is0186: clau|AEV94470.1 ldam|AH70_06485 pent|YP_803680
is0187: clau|AEV94471.1 ldam|AH70_06490 pent|YP_803681
is0188: clau|AEV94472.1 ldam|AH70_06495 pent|YP_803682
is0189: clau|AEV94473.1 ldam|AH70_06500 pent|YP_803683
is0190: clau|AEV94474.1 ldam|AH70_06505 pent|YP_803684
is0191: clau|AEV94475.1 ldam|AH70_06510 pent|YP_803685
is0192: clau|AEV94477.1 ldam|AH70_06515 pent|YP_803686
is0193: clau|AEV94478.1 ldam|AH70_06520 pent|YP_803687
is0194: clau|AEV94483.1 ldam|AH70_04900 pent|YP_803827
is0195: clau|AEV94485.1 ldam|AH70_07620 pent|YP_803829
is0196: clau|AEV94486.1 ldam|AH70_07615 pent|YP_803830
is0197: clau|AEV94487.1 ldam|AH70_07590 pent|YP_803831
is0198: clau|AEV94488.1 ldam|AH70_07585 pent|YP_803832
is0199: clau|AEV94489.1 ldam|AH70_07580 pent|YP_803834
is0200: clau|AEV94493.1 ldam|AH70_07555 pent|YP_803838
is0201: clau|AEV94494.1 ldam|AH70_07560 pent|YP_803837
is0202: clau|AEV94495.1 pent|YP_803836 ldam|AH70_07570
is0203: clau|AEV94497.1 ldam|AH70_07450 pent|YP_803849
is0204: clau|AEV94499.1 ldam|AH70_10575 pent|YP_803848
is0205: clau|AEV94506.1 ldam|AH70_07220 pent|YP_805264
is0206: clau|AEV94509.1 ldam|AH70_07480 pent|YP_803845
is0207: clau|AEV94511.1 ldam|AH70_06820 pent|YP_805059
is0208: clau|AEV94513.1 ldam|AH70_06810 pent|YP_805057
is0209: clau|AEV94514.1 ldam|AH70_10960 pent|YP_805056
is0210: clau|AEV94516.1 ldam|AH70_10895 pent|YP_805053
is0211: clau|AEV94517.1 ldam|AH70_10900 pent|YP_805052
is0212: clau|AEV94518.1 ldam|AH70_10905 pent|YP_805051
is0213: clau|AEV94519.1 ldam|AH70_10910 pent|YP_805050
is0214: clau|AEV94522.1 ldam|AH70_01095 pent|YP_805049
is0215: clau|AEV94525.1 ldam|AH70_01075 pent|YP_805048
is0216: clau|AEV94526.1 ldam|AH70_08220 pent|YP_805046
is0217: clau|AEV94527.1 ldam|AH70_09645 pent|YP_805045
is0218: clau|AEV94528.1 ldam|AH70_07790 pent|YP_805044
is0219: clau|AEV94529.1 ldam|AH70_07155 pent|YP_805043
is0220: clau|AEV94531.1 ldam|AH70_07800 pent|YP_805041
is0221: clau|AEV94532.1 ldam|AH70_06010 pent|YP_803765
is0222: clau|AEV94536.1 pent|YP_805039 ldam|AH70_07855
is0223: clau|AEV94537.1 ldam|AH70_07865 pent|YP_805037
is0224: clau|AEV94538.1 ldam|AH70_07870 pent|YP_805036
is0225: clau|AEV94539.1 ldam|AH70_07875 pent|YP_805035
is0226: clau|AEV94540.1 ldam|AH70_07880 pent|YP_805034
is0227: clau|AEV94541.1 ldam|AH70_07885 pent|YP_805033
is0228: clau|AEV94542.1 ldam|AH70_07890 pent|YP_805032
is0229: clau|AEV94543.1 ldam|AH70_07895 pent|YP_805031
is0230: clau|AEV94544.1 ldam|AH70_07900 pent|YP_805030
is0231: clau|AEV94545.1 ldam|AH70_07905 pent|YP_805029

is0232: clau|AEV94546.1 ldam|AH70_07915 pent|YP_805028
is0233: clau|AEV94547.1 ldam|AH70_07920 pent|YP_805027
is0234: clau|AEV94548.1 ldam|AH70_07925 pent|YP_805026
is0235: clau|AEV94549.1 ldam|AH70_07930 pent|YP_805025
is0236: clau|AEV94550.1 ldam|AH70_07935 pent|YP_805024
is0237: clau|AEV94551.1 ldam|AH70_07940 pent|YP_805023
is0238: clau|AEV94552.1 ldam|AH70_07945 pent|YP_805022
is0239: clau|AEV94553.1 ldam|AH70_07950 pent|YP_805021
is0240: clau|AEV94554.1 ldam|AH70_07955 pent|YP_805020
is0241: clau|AEV94556.1 ldam|AH70_07980 pent|YP_805019
is0242: clau|AEV94557.1 ldam|AH70_07985 pent|YP_805018
is0243: clau|AEV94558.1 ldam|AH70_07990 pent|YP_805017
is0244: clau|AEV94562.1 ldam|AH70_07390 pent|YP_804022
is0245: clau|AEV94563.1 ldam|AH70_07385 pent|YP_804023
is0246: clau|AEV94564.1 ldam|AH70_07380 pent|YP_804024
is0247: clau|AEV94565.1 ldam|AH70_07375 pent|YP_804025
is0248: clau|AEV94566.1 ldam|AH70_07370 pent|YP_804026
is0249: clau|AEV94588.1 ldam|AH70_07995 pent|YP_805016
is0250: clau|AEV94589.1 ldam|AH70_08000 pent|YP_805015
is0251: clau|AEV94590.1 ldam|AH70_08005 pent|YP_805014
is0252: clau|AEV94591.1 ldam|AH70_10055 pent|YP_805013
is0253: clau|AEV94592.1 ldam|AH70_10050 pent|YP_805012
is0254: clau|AEV94593.1 ldam|AH70_10045 pent|YP_805011
is0255: clau|AEV94594.1 ldam|AH70_10040 pent|YP_805010
is0256: clau|AEV94595.1 ldam|AH70_10035 pent|YP_805009
is0257: clau|AEV94596.1 ldam|AH70_10030 pent|YP_805008
is0258: clau|AEV94597.1 ldam|AH70_10025 pent|YP_805007
is0259: clau|AEV94598.1 ldam|AH70_10020 pent|YP_805006
is0260: clau|AEV94599.1 ldam|AH70_10015 pent|YP_805005
is0261: clau|AEV94600.1 ldam|AH70_10840 pent|YP_805004
is0262: clau|AEV94605.1 ldam|AH70_03345 pent|YP_805082
is0263: clau|AEV94609.1 ldam|AH70_01490 pent|YP_805141
is0264: clau|AEV94614.1 ldam|AH70_02540 pent|YP_803582
is0265: clau|AEV94621.1 ldam|AH70_10800 pent|YP_805001
is0266: clau|AEV94622.1 ldam|AH70_10450 pent|YP_805000
is0267: clau|AEV94623.1 ldam|AH70_10455 pent|YP_804999
is0268: clau|AEV94625.1 ldam|AH70_05540 pent|YP_805164
is0269: clau|AEV94629.1 ldam|AH70_10460 pent|YP_804998
is0270: clau|AEV94630.1 ldam|AH70_10465 pent|YP_804997
is0271: clau|AEV94631.1 ldam|AH70_10470 pent|YP_804996
is0272: clau|AEV94641.1 ldam|AH70_10485 pent|YP_804993
is0273: clau|AEV94642.1 ldam|AH70_10490 pent|YP_804992
is0274: clau|AEV94651.1 ldam|AH70_10510 pent|YP_804991
is0275: clau|AEV94652.1 ldam|AH70_10515 pent|YP_804990
is0276: clau|AEV94653.1 ldam|AH70_10520 pent|YP_804989
is0277: clau|AEV94654.1 ldam|AH70_10525 pent|YP_804988
is0278: clau|AEV94655.1 ldam|AH70_10505 pent|YP_804987
is0279: clau|AEV94657.1 ldam|AH70_10540 pent|YP_804986
is0280: clau|AEV94660.1 ldam|AH70_03190 pent|YP_803634
is0281: clau|AEV94673.1 ldam|AH70_10545 pent|YP_804984
is0282: clau|AEV94678.1 ldam|AH70_10585 pent|YP_804983
is0283: clau|AEV94680.1 ldam|AH70_10590 pent|YP_804980
is0284: clau|AEV94681.1 ldam|AH70_10595 pent|YP_804979
is0285: clau|AEV94682.1 ldam|AH70_10600 pent|YP_804978
is0286: clau|AEV94683.1 ldam|AH70_10605 pent|YP_804977
is0287: clau|AEV94684.1 ldam|AH70_10610 pent|YP_804976
is0288: clau|AEV94685.1 ldam|AH70_10615 pent|YP_804975
is0289: clau|AEV94686.1 ldam|AH70_10620 pent|YP_804974
is0290: clau|AEV94688.1 ldam|AH70_09585 pent|YP_803876
is0291: clau|AEV94690.1 ldam|AH70_09330 pent|YP_803878
is0292: clau|AEV94693.1 ldam|AH70_09295 pent|YP_803879
is0293: clau|AEV94695.1 ldam|AH70_09290 pent|YP_803880
is0294: clau|AEV94696.1 ldam|AH70_01950 pent|YP_803883
is0295: clau|AEV94697.1 ldam|AH70_01955 pent|YP_803882
is0296: clau|AEV94698.1 ldam|AH70_01970 pent|YP_803881
is0297: clau|AEV94700.1 ldam|AH70_01945 pent|YP_803884
is0298: clau|AEV94701.1 ldam|AH70_01910 pent|YP_803885
is0299: clau|AEV94704.1 ldam|AH70_01705 pent|YP_805254
is0300: clau|AEV94711.1 ldam|AH70_01480 pent|YP_805226
is0301: clau|AEV94719.1 ldam|AH70_01110 pent|YP_803769
is0302: clau|AEV94723.1 ldam|AH70_09640 pent|YP_805077
is0303: clau|AEV94728.1 ldam|AH70_05310 pent|YP_805076
is0304: clau|AEV94729.1 ldam|AH70_05305 pent|YP_805075
is0305: clau|AEV94730.1 ldam|AH70_09710 pent|YP_805074
is0306: clau|AEV94731.1 ldam|AH70_08350 pent|YP_805073
is0307: clau|AEV94732.1 ldam|AH70_02125 pent|YP_805070
is0308: clau|AEV94733.1 ldam|AH70_02130 pent|YP_805069
is0309: clau|AEV94741.1 ldam|AH70_07545 pent|YP_803839
is0310: clau|AEV94742.1 ldam|AH70_07540 pent|YP_803840
is0311: clau|AEV94743.1 ldam|AH70_07535 pent|YP_803841
is0312: clau|AEV94746.1 ldam|AH70_07510 pent|YP_803843
is0313: clau|AEV94747.1 ldam|AH70_07505 pent|YP_803844
is0314: clau|AEV94750.1 ldam|AH70_10725 pent|YP_803863
is0315: clau|AEV94752.1 ldam|AH70_06440 pent|YP_803868
is0316: clau|AEV94756.1 ldam|AH70_06415 pent|YP_803873
is0317: clau|AEV94757.1 ldam|AH70_04265 pent|YP_803671
is0318: clau|AEV94758.1 ldam|AH70_04260 pent|YP_803670
is0319: clau|AEV94759.1 ldam|AH70_02330 pent|YP_805212
is0320: clau|AEV94764.1 ldam|AH70_09605 pent|YP_803874
is0321: clau|AEV94765.1 ldam|AH70_09600 pent|YP_803875
is0322: clau|AEV94768.1 ldam|AH70_10625 pent|YP_804973
is0323: clau|AEV94769.1 ldam|AH70_10630 pent|YP_804972
is0324: clau|AEV94770.1 ldam|AH70_10635 pent|YP_804971
is0325: clau|AEV94771.1 ldam|AH70_10640 pent|YP_804970
is0326: clau|AEV94772.1 ldam|AH70_04750 pent|YP_804968
is0327: clau|AEV94773.1 ldam|AH70_04755 pent|YP_804967
is0328: clau|AEV94774.1 ldam|AH70_04760 pent|YP_804966
is0329: clau|AEV94775.1 ldam|AH70_04765 pent|YP_804965
is0330: clau|AEV94776.1 ldam|AH70_04770 pent|YP_804964
is0331: clau|AEV94777.1 ldam|AH70_04775 pent|YP_804963
is0332: clau|AEV94779.1 ldam|AH70_04170 pent|YP_805138
is0333: clau|AEV94781.1 ldam|AH70_07185 pent|YP_803851
is0334: clau|AEV94782.1 ldam|AH70_07180 pent|YP_803852
is0335: clau|AEV94783.1 ldam|AH70_07175 pent|YP_803853
is0336: clau|AEV94784.1 ldam|AH70_07170 pent|YP_803854
is0337: clau|AEV94786.1 ldam|AH70_06525 pent|YP_805147
is0338: clau|AEV94791.1 ldam|AH70_04795 pent|YP_804962
is0339: clau|AEV94793.1 ldam|AH70_04810 pent|YP_804960
is0340: clau|AEV94797.1 ldam|AH70_04820 pent|YP_804958
is0341: clau|AEV94798.1 ldam|AH70_04825 pent|YP_804957
is0342: clau|AEV94799.1 ldam|AH70_04830 pent|YP_804956
is0343: clau|AEV94800.1 ldam|AH70_04835 pent|YP_804955
is0344: clau|AEV94801.1 ldam|AH70_04840 pent|YP_804954
is0345: clau|AEV94803.1 ldam|AH70_04850 pent|YP_804952
is0346: clau|AEV94804.1 ldam|AH70_04855 pent|YP_804951
is0347: clau|AEV94805.1 ldam|AH70_04860 pent|YP_804950
is0348: clau|AEV94806.1 ldam|AH70_04865 pent|YP_804949
is0349: clau|AEV94808.1 ldam|AH70_04875 pent|YP_804947
is0350: clau|AEV94809.1 ldam|AH70_04880 pent|YP_804946
is0351: clau|AEV94810.1 ldam|AH70_04885 pent|YP_804945
is0352: clau|AEV94811.1 ldam|AH70_04890 pent|YP_804944
is0353: clau|AEV94812.1 ldam|AH70_01730 pent|YP_803957
is0354: clau|AEV94813.1 ldam|AH70_01735 pent|YP_803958
is0355: clau|AEV94814.1 ldam|AH70_01740 pent|YP_803959
is0356: clau|AEV94815.1 ldam|AH70_09845 pent|YP_803960
is0357: clau|AEV94816.1 ldam|AH70_09850 pent|YP_803961
is0358: clau|AEV94817.1 ldam|AH70_09855 pent|YP_803962
is0359: clau|AEV94818.1 ldam|AH70_09860 pent|YP_803963
is0360: clau|AEV94819.1 ldam|AH70_09865 pent|YP_803964
is0361: clau|AEV94820.1 ldam|AH70_09870 pent|YP_803965
is0362: clau|AEV94822.1 ldam|AH70_09880 pent|YP_803968
is0363: clau|AEV94823.1 ldam|AH70_09885 pent|YP_803969
is0364: clau|AEV94824.1 ldam|AH70_09890 pent|YP_803970
is0365: clau|AEV94825.1 ldam|AH70_09895 pent|YP_803971
is0366: clau|AEV94826.1 ldam|AH70_09900 pent|YP_803972
is0367: clau|AEV94827.1 ldam|AH70_09905 pent|YP_803973
is0368: clau|AEV94828.1 ldam|AH70_09910 pent|YP_803974
is0369: clau|AEV94829.1 ldam|AH70_09915 pent|YP_803975
is0370: clau|AEV94830.1 ldam|AH70_09920 pent|YP_803976
is0371: clau|AEV94831.1 ldam|AH70_09925 pent|YP_803977
is0372: clau|AEV94833.1 ldam|AH70_09935 pent|YP_803979
is0373: clau|AEV94834.1 ldam|AH70_09940 pent|YP_803980
is0374: clau|AEV94835.1 ldam|AH70_09945 pent|YP_803981
is0375: clau|AEV94836.1 ldam|AH70_09950 pent|YP_803982
is0376: clau|AEV94837.1 ldam|AH70_09955 pent|YP_803983
is0377: clau|AEV94839.1 ldam|AH70_09960 pent|YP_803984
is0378: clau|AEV94840.1 ldam|AH70_09970 pent|YP_803986
is0379: clau|AEV94841.1 ldam|AH70_04645 pent|YP_803987
is0380: clau|AEV94842.1 ldam|AH70_09985 pent|YP_803988
is0381: clau|AEV94843.1 ldam|AH70_09990 pent|YP_803989
is0382: clau|AEV94844.1 ldam|AH70_09995 pent|YP_803990
is0383: clau|AEV94845.1 ldam|AH70_07740 pent|YP_803993
is0384: clau|AEV94846.1 ldam|AH70_07730 pent|YP_803994
is0385: clau|AEV94847.1 ldam|AH70_07725 pent|YP_803995
is0386: clau|AEV94848.1 ldam|AH70_07720 pent|YP_803996
is0387: clau|AEV94849.1 ldam|AH70_07715 pent|YP_803997
is0388: clau|AEV94850.1 ldam|AH70_07710 pent|YP_803998
is0389: clau|AEV94851.1 ldam|AH70_07705 pent|YP_803999
is0390: clau|AEV94853.1 ldam|AH70_07690 pent|YP_804000
is0391: clau|AEV94854.1 ldam|AH70_07685 pent|YP_804001
is0392: clau|AEV94855.1 ldam|AH70_07680 pent|YP_804002
is0393: clau|AEV94861.1 ldam|AH70_07670 pent|YP_804003
is0394: clau|AEV94862.1 ldam|AH70_07665 pent|YP_804004
is0395: clau|AEV94863.1 ldam|AH70_07660 pent|YP_804005
is0396: clau|AEV94864.1 ldam|AH70_07655 pent|YP_804006
is0397: clau|AEV94865.1 ldam|AH70_07650 pent|YP_804007

is0398: clau|AEV94866.1 ldam|AH70_07645 pent|YP_804008
is0399: clau|AEV94867.1 ldam|AH70_07640 pent|YP_804009
is0400: clau|AEV94868.1 ldam|AH70_07635 pent|YP_804010
is0401: clau|AEV94869.1 ldam|AH70_07630 pent|YP_804011
is0402: clau|AEV94870.1 ldam|AH70_03380 pent|YP_804012
is0403: clau|AEV94871.1 ldam|AH70_03375 pent|YP_804013
is0404: clau|AEV94872.1 ldam|AH70_03370 pent|YP_804014
is0405: clau|AEV94873.1 ldam|AH70_03365 pent|YP_804015
is0406: clau|AEV94874.1 ldam|AH70_03355 pent|YP_804016
is0407: clau|AEV94882.1 ldam|AH70_03305 pent|YP_804073
is0408: clau|AEV94895.1 ldam|AH70_09455 pent|YP_804503
is0409: clau|AEV94897.1 ldam|AH70_09445 pent|YP_804501
is0410: clau|AEV94898.1 ldam|AH70_09435 pent|YP_804499
is0411: clau|AEV94899.1 ldam|AH70_09430 pent|YP_804498
is0412: clau|AEV94907.1 ldam|AH70_03250 pent|YP_804835
is0413: clau|AEV94908.1 ldam|AH70_03245 pent|YP_804834
is0414: clau|AEV94911.1 ldam|AH70_03240 pent|YP_804833
is0415: clau|AEV94912.1 ldam|AH70_03235 pent|YP_804832
is0416: clau|AEV94923.1 ldam|AH70_03230 pent|YP_804831
is0417: clau|AEV94924.1 ldam|AH70_07405 pent|YP_804020
is0418: clau|AEV94925.1 ldam|AH70_07400 pent|YP_804021
is0419: clau|AEV94926.1 ldam|AH70_08595 pent|YP_803643
is0420: clau|AEV94928.1 ldam|AH70_01805 pent|YP_804102
is0421: clau|AEV94934.1 ldam|AH70_05290 pent|YP_803763
is0422: clau|AEV94937.1 ldam|AH70_03200 pent|YP_804822
is0423: clau|AEV94938.1 ldam|AH70_03160 pent|YP_804821
is0424: clau|AEV94939.1 ldam|AH70_03155 pent|YP_804820
is0425: clau|AEV94940.1 ldam|AH70_00760 pent|YP_804819
is0426: clau|AEV94941.1 ldam|AH70_00765 pent|YP_804818
is0427: clau|AEV94942.1 ldam|AH70_00770 pent|YP_804817
is0428: clau|AEV94943.1 ldam|AH70_01885 pent|YP_803767
is0429: clau|AEV94944.1 ldam|AH70_01880 pent|YP_803768
is0430: clau|AEV94945.1 ldam|AH70_00775 pent|YP_804816
is0431: clau|AEV94946.1 ldam|AH70_00780 pent|YP_804815
is0432: clau|AEV94947.1 ldam|AH70_00785 pent|YP_804814
is0433: clau|AEV94948.1 ldam|AH70_00790 pent|YP_804813
is0434: clau|AEV94949.1 ldam|AH70_00795 pent|YP_804812
is0435: clau|AEV94950.1 ldam|AH70_00800 pent|YP_804811
is0436: clau|AEV94951.1 ldam|AH70_00805 pent|YP_804810
is0437: clau|AEV94952.1 ldam|AH70_00810 pent|YP_804809
is0438: clau|AEV94953.1 ldam|AH70_02280 pent|YP_804806
is0439: clau|AEV94954.1 ldam|AH70_02285 pent|YP_804805
is0440: clau|AEV94955.1 ldam|AH70_02290 pent|YP_804804
is0441: clau|AEV94956.1 ldam|AH70_02295 pent|YP_804803
is0442: clau|AEV94957.1 ldam|AH70_02300 pent|YP_804802
is0443: clau|AEV94958.1 ldam|AH70_02305 pent|YP_804801
is0444: clau|AEV94959.1 ldam|AH70_02310 pent|YP_804800
is0445: clau|AEV94960.1 ldam|AH70_02315 pent|YP_804799
is0446: clau|AEV94961.1 ldam|AH70_06790 pent|YP_804798
is0447: clau|AEV94963.1 ldam|AH70_06775 pent|YP_804796
is0448: clau|AEV94964.1 ldam|AH70_06770 pent|YP_804795
is0449: clau|AEV94965.1 ldam|AH70_06765 pent|YP_804794
is0450: clau|AEV94966.1 ldam|AH70_06760 pent|YP_804793
is0451: clau|AEV94967.1 ldam|AH70_06755 pent|YP_804792
is0452: clau|AEV94969.1 ldam|AH70_06735 pent|YP_804791
is0453: clau|AEV94970.1 ldam|AH70_06730 pent|YP_804790
is0454: clau|AEV94971.1 ldam|AH70_06725 pent|YP_804789
is0455: clau|AEV94972.1 ldam|AH70_06720 pent|YP_804788
is0456: clau|AEV94974.1 ldam|AH70_06715 pent|YP_804786
is0457: clau|AEV94975.1 ldam|AH70_06710 pent|YP_805229
is0458: clau|AEV94976.1 ldam|AH70_06705 pent|YP_804785
is0459: clau|AEV94977.1 ldam|AH70_06695 pent|YP_804784
is0460: clau|AEV94978.1 ldam|AH70_06690 pent|YP_804783
is0461: clau|AEV94979.1 ldam|AH70_06680 pent|YP_804781
is0462: clau|AEV94980.1 ldam|AH70_06665 pent|YP_804780
is0463: clau|AEV94981.1 ldam|AH70_06660 pent|YP_804779
is0464: clau|AEV94982.1 ldam|AH70_06655 pent|YP_804778
is0465: clau|AEV94983.1 ldam|AH70_06650 pent|YP_804777
is0466: clau|AEV94984.1 ldam|AH70_06645 pent|YP_804776
is0467: clau|AEV94985.1 ldam|AH70_06640 pent|YP_804775
is0468: clau|AEV94986.1 ldam|AH70_06635 pent|YP_804774
is0469: clau|AEV94987.1 ldam|AH70_06630 pent|YP_804773
is0470: clau|AEV94988.1 ldam|AH70_06625 pent|YP_804772
is0471: clau|AEV94989.1 ldam|AH70_06620 pent|YP_804771
is0472: clau|AEV94990.1 ldam|AH70_06615 pent|YP_804770
is0473: clau|AEV94991.1 ldam|AH70_03920 pent|YP_804769
is0474: clau|AEV94992.1 ldam|AH70_03925 pent|YP_804768
is0475: clau|AEV94993.1 ldam|AH70_03930 pent|YP_804767
is0476: clau|AEV94995.1 ldam|AH70_03935 pent|YP_804765
is0477: clau|AEV94996.1 ldam|AH70_03940 pent|YP_804764
is0478: clau|AEV94997.1 ldam|AH70_03945 pent|YP_804763
is0479: clau|AEV94999.1 ldam|AH70_03955 pent|YP_804761
is0480: clau|AEV95000.1 ldam|AH70_03960 pent|YP_804760
is0481: clau|AEV95001.1 ldam|AH70_03965 pent|YP_804759
is0482: clau|AEV95002.1 ldam|AH70_03970 pent|YP_804758
is0483: clau|AEV95003.1 ldam|AH70_03975 pent|YP_804757
is0484: clau|AEV95004.1 ldam|AH70_03980 pent|YP_804756
is0485: clau|AEV95005.1 ldam|AH70_03985 pent|YP_804755
is0486: clau|AEV95006.1 ldam|AH70_03990 pent|YP_804754
is0487: clau|AEV95008.1 ldam|AH70_03995 pent|YP_804752
is0488: clau|AEV95009.1 ldam|AH70_04000 pent|YP_804751
is0489: clau|AEV95010.1 ldam|AH70_04005 pent|YP_804750
is0490: clau|AEV95011.1 ldam|AH70_04010 pent|YP_804749
is0491: clau|AEV95012.1 ldam|AH70_10850 pent|YP_804748
is0492: clau|AEV95013.1 ldam|AH70_10855 pent|YP_804747
is0493: clau|AEV95014.1 ldam|AH70_10860 pent|YP_804746
is0494: clau|AEV95015.1 ldam|AH70_10865 pent|YP_804745
is0495: clau|AEV95016.1 ldam|AH70_10870 pent|YP_804744
is0496: clau|AEV95017.1 ldam|AH70_10875 pent|YP_804743
is0497: clau|AEV95018.1 ldam|AH70_10880 pent|YP_804742
is0498: clau|AEV95019.1 ldam|AH70_10885 pent|YP_804741
is0499: clau|AEV95020.1 ldam|AH70_02010 pent|YP_804740
is0500: clau|AEV95021.1 ldam|AH70_02005 pent|YP_804739
is0501: clau|AEV95022.1 ldam|AH70_02000 pent|YP_804738
is0502: clau|AEV95024.1 ldam|AH70_01990 pent|YP_804736
is0503: clau|AEV95026.1 ldam|AH70_08375 pent|YP_804734
is0504: clau|AEV95027.1 ldam|AH70_08380 pent|YP_804733
is0505: clau|AEV95031.1 ldam|AH70_03215 pent|YP_804826
is0506: clau|AEV95032.1 ldam|AH70_10175 pent|YP_803590
is0507: clau|AEV95033.1 ldam|AH70_10720 pent|YP_804727
is0508: clau|AEV95035.1 ldam|AH70_02200 pent|YP_804723
is0509: clau|AEV95037.1 ldam|AH70_02205 pent|YP_804721
is0510: clau|AEV95038.1 ldam|AH70_02210 pent|YP_804720
is0511: clau|AEV95039.1 ldam|AH70_02215 pent|YP_804719
is0512: clau|AEV95040.1 ldam|AH70_02220 pent|YP_804718
is0513: clau|AEV95042.1 ldam|AH70_02230 pent|YP_804716
is0514: clau|AEV95044.1 ldam|AH70_02240 pent|YP_804714
is0515: clau|AEV95045.1 ldam|AH70_02245 pent|YP_804713
is0516: clau|AEV95046.1 ldam|AH70_02250 pent|YP_804712
is0517: clau|AEV95047.1 ldam|AH70_02255 pent|YP_804711
is0518: clau|AEV95048.1 ldam|AH70_02260 pent|YP_804710
is0519: clau|AEV95049.1 ldam|AH70_02265 pent|YP_804709
is0520: clau|AEV95050.1 ldam|AH70_02270 pent|YP_804708
is0521: clau|AEV95051.1 ldam|AH70_02275 pent|YP_804707
is0522: clau|AEV95054.1 ldam|AH70_08985 pent|YP_804705
is0523: clau|AEV95055.1 ldam|AH70_08990 pent|YP_804704
is0524: clau|AEV95056.1 ldam|AH70_08995 pent|YP_804703
is0525: clau|AEV95057.1 ldam|AH70_09000 pent|YP_804702
is0526: clau|AEV95060.1 ldam|AH70_09015 pent|YP_804700
is0527: clau|AEV95061.1 ldam|AH70_09020 pent|YP_804699
is0528: clau|AEV95062.1 ldam|AH70_09025 pent|YP_804698
is0529: clau|AEV95063.1 ldam|AH70_09030 pent|YP_804697
is0530: clau|AEV95064.1 ldam|AH70_09035 pent|YP_804696
is0531: clau|AEV95065.1 ldam|AH70_09315 pent|YP_804730
is0532: clau|AEV95066.1 ldam|AH70_09310 pent|YP_804729
is0533: clau|AEV95067.1 ldam|AH70_09040 pent|YP_804695
is0534: clau|AEV95068.1 ldam|AH70_09045 pent|YP_804694
is0535: clau|AEV95069.1 ldam|AH70_09050 pent|YP_804693
is0536: clau|AEV95070.1 ldam|AH70_09055 pent|YP_804692
is0537: clau|AEV95071.1 ldam|AH70_09065 pent|YP_804691
is0538: clau|AEV95072.1 ldam|AH70_09070 pent|YP_804690
is0539: clau|AEV95073.1 ldam|AH70_09080 pent|YP_804688
is0540: clau|AEV95074.1 ldam|AH70_09085 pent|YP_804687
is0541: clau|AEV95075.1 ldam|AH70_09090 pent|YP_804686
is0542: clau|AEV95076.1 ldam|AH70_09095 pent|YP_804685
is0543: clau|AEV95077.1 ldam|AH70_09100 pent|YP_804684
is0544: clau|AEV95078.1 ldam|AH70_09105 pent|YP_804683
is0545: clau|AEV95079.1 ldam|AH70_09110 pent|YP_804682
is0546: clau|AEV95080.1 ldam|AH70_09115 pent|YP_804681
is0547: clau|AEV95081.1 ldam|AH70_09120 pent|YP_804680
is0548: clau|AEV95082.1 ldam|AH70_09125 pent|YP_804679
is0549: clau|AEV95083.1 ldam|AH70_09135 pent|YP_804678
is0550: clau|AEV95084.1 ldam|AH70_09140 pent|YP_804677
is0551: clau|AEV95085.1 ldam|AH70_09145 pent|YP_804676
is0552: clau|AEV95086.1 ldam|AH70_09150 pent|YP_804675
is0553: clau|AEV95088.1 ldam|AH70_09160 pent|YP_804673
is0554: clau|AEV95089.1 ldam|AH70_09165 pent|YP_804672
is0555: clau|AEV95090.1 ldam|AH70_09170 pent|YP_804671
is0556: clau|AEV95091.1 ldam|AH70_09175 pent|YP_804670
is0557: clau|AEV95092.1 ldam|AH70_09180 pent|YP_804669
is0558: clau|AEV95093.1 ldam|AH70_09185 pent|YP_804668
is0559: clau|AEV95094.1 ldam|AH70_09190 pent|YP_804667
is0560: clau|AEV95095.1 ldam|AH70_09195 pent|YP_804666
is0561: clau|AEV95096.1 ldam|AH70_09200 pent|YP_804665
is0562: clau|AEV95097.1 ldam|AH70_09205 pent|YP_804664
is0563: clau|AEV95098.1 ldam|AH70_09210 pent|YP_804663

is0564: clau|AEV95099.1 ldam|AH70_09215 pent|YP_804662
is0565: clau|AEV95100.1 ldam|AH70_09220 pent|YP_804661
is0566: clau|AEV95101.1 ldam|AH70_09225 pent|YP_804660
is0567: clau|AEV95103.1 ldam|AH70_01770 pent|YP_804659
is0568: clau|AEV95104.1 ldam|AH70_08845 pent|YP_804658
is0569: clau|AEV95105.1 ldam|AH70_01775 pent|YP_804657
is0570: clau|AEV95106.1 ldam|AH70_01780 pent|YP_804656
is0571: clau|AEV95107.1 ldam|AH70_01785 pent|YP_804655
is0572: clau|AEV95108.1 ldam|AH70_01790 pent|YP_804654
is0573: clau|AEV95109.1 ldam|AH70_01795 pent|YP_804653
is0574: clau|AEV95110.1 ldam|AH70_01800 pent|YP_804652
is0575: clau|AEV95111.1 ldam|AH70_00400 pent|YP_804651
is0576: clau|AEV95112.1 ldam|AH70_00415 pent|YP_804650
is0577: clau|AEV95114.1 ldam|AH70_00425 pent|YP_804648
is0578: clau|AEV95116.1 ldam|AH70_00435 pent|YP_804647
is0579: clau|AEV95117.1 ldam|AH70_00440 pent|YP_804646
is0580: clau|AEV95118.1 ldam|AH70_00445 pent|YP_804645
is0581: clau|AEV95119.1 ldam|AH70_00450 pent|YP_804644
is0582: clau|AEV95120.1 ldam|AH70_00455 pent|YP_804643
is0583: clau|AEV95121.1 ldam|AH70_00460 pent|YP_804642
is0584: clau|AEV95122.1 ldam|AH70_00465 pent|YP_804641
is0585: clau|AEV95123.1 ldam|AH70_00470 pent|YP_804640
is0586: clau|AEV95124.1 ldam|AH70_00475 pent|YP_804639
is0587: clau|AEV95125.1 ldam|AH70_00480 pent|YP_804638
is0588: clau|AEV95126.1 ldam|AH70_00485 pent|YP_804637
is0589: clau|AEV95127.1 ldam|AH70_00490 pent|YP_804636
is0590: clau|AEV95128.1 ldam|AH70_00495 pent|YP_804635
is0591: clau|AEV95130.1 ldam|AH70_00505 pent|YP_804633
is0592: clau|AEV95131.1 ldam|AH70_00510 pent|YP_804632
is0593: clau|AEV95132.1 ldam|AH70_00515 pent|YP_804631
is0594: clau|AEV95133.1 ldam|AH70_00520 pent|YP_804630
is0595: clau|AEV95134.1 ldam|AH70_00525 pent|YP_805203
is0596: clau|AEV95136.1 ldam|AH70_07305 pent|YP_804628
is0597: clau|AEV95137.1 ldam|AH70_07300 pent|YP_804627
is0598: clau|AEV95138.1 ldam|AH70_07290 pent|YP_804626
is0599: clau|AEV95139.1 ldam|AH70_07285 pent|YP_804625
is0600: clau|AEV95157.1 ldam|AH70_05010 pent|YP_804596
is0601: clau|AEV95158.1 ldam|AH70_05015 pent|YP_804595
is0602: clau|AEV95159.1 ldam|AH70_05020 pent|YP_804594
is0603: clau|AEV95160.1 ldam|AH70_05030 pent|YP_804592
is0604: clau|AEV95161.1 ldam|AH70_05035 pent|YP_804591
is0605: clau|AEV95162.1 ldam|AH70_05040 pent|YP_804590
is0606: clau|AEV95163.1 ldam|AH70_05045 pent|YP_804589
is0607: clau|AEV95164.1 ldam|AH70_05050 pent|YP_804587
is0608: clau|AEV95165.1 ldam|AH70_05055 pent|YP_804586
is0609: clau|AEV95166.1 ldam|AH70_05060 pent|YP_804585
is0610: clau|AEV95167.1 ldam|AH70_05065 pent|YP_804584
is0611: clau|AEV95168.1 ldam|AH70_05070 pent|YP_804583
is0612: clau|AEV95169.1 ldam|AH70_05075 pent|YP_804582
is0613: clau|AEV95170.1 ldam|AH70_05080 pent|YP_804581
is0614: clau|AEV95171.1 ldam|AH70_05085 pent|YP_804580
is0615: clau|AEV95172.1 ldam|AH70_05090 pent|YP_804579
is0616: clau|AEV95174.1 ldam|AH70_05100 pent|YP_804577
is0617: clau|AEV95175.1 ldam|AH70_05105 pent|YP_804576
is0618: clau|AEV95177.1 ldam|AH70_05115 pent|YP_804575
is0619: clau|AEV95178.1 ldam|AH70_05120 pent|YP_804574
is0620: clau|AEV95179.1 ldam|AH70_05125 pent|YP_804573
is0621: clau|AEV95180.1 ldam|AH70_05130 pent|YP_804572
is0622: clau|AEV95181.1 ldam|AH70_05135 pent|YP_804571
is0623: clau|AEV95182.1 ldam|AH70_05140 pent|YP_804570
is0624: clau|AEV95183.1 ldam|AH70_05145 pent|YP_804569
is0625: clau|AEV95184.1 ldam|AH70_05150 pent|YP_804568
is0626: clau|AEV95185.1 ldam|AH70_05155 pent|YP_804567
is0627: clau|AEV95186.1 ldam|AH70_05160 pent|YP_804564
is0628: clau|AEV95187.1 ldam|AH70_05175 pent|YP_804561
is0629: clau|AEV95188.1 ldam|AH70_05180 pent|YP_804560
is0630: clau|AEV95189.1 ldam|AH70_05165 pent|YP_804563
is0631: clau|AEV95190.1 ldam|AH70_05170 pent|YP_804562
is0632: clau|AEV95191.1 ldam|AH70_05195 pent|YP_804470
is0633: clau|AEV95192.1 ldam|AH70_05200 pent|YP_804469
is0634: clau|AEV95193.1 ldam|AH70_05205 pent|YP_804468
is0635: clau|AEV95194.1 ldam|AH70_05210 pent|YP_804463
is0636: clau|AEV95195.1 ldam|AH70_05215 pent|YP_804462
is0637: clau|AEV95197.1 ldam|AH70_05220 pent|YP_804460
is0638: clau|AEV95198.1 ldam|AH70_05225 pent|YP_804459
is0639: clau|AEV95199.1 ldam|AH70_05230 pent|YP_804458
is0640: clau|AEV95201.1 ldam|AH70_05240 pent|YP_804456
is0641: clau|AEV95202.1 ldam|AH70_05245 pent|YP_804455
is0642: clau|AEV95203.1 ldam|AH70_05250 pent|YP_804454
is0643: clau|AEV95204.1 ldam|AH70_10925 pent|YP_804452
is0644: clau|AEV95205.1 ldam|AH70_02775 pent|YP_804451
is0645: clau|AEV95206.1 ldam|AH70_07345 pent|YP_804450
is0646: clau|AEV95211.1 ldam|AH70_05760 pent|YP_804447
is0647: clau|AEV95212.1 ldam|AH70_08850 pent|YP_804446
is0648: clau|AEV95214.1 ldam|AH70_08855 pent|YP_804444
is0649: clau|AEV95215.1 ldam|AH70_05000 pent|YP_804598
is0650: clau|AEV95216.1 ldam|AH70_04995 pent|YP_804599
is0651: clau|AEV95219.1 ldam|AH70_04980 pent|YP_804602
is0652: clau|AEV95220.1 ldam|AH70_04975 pent|YP_804603
is0653: clau|AEV95221.1 ldam|AH70_04970 pent|YP_804604
is0654: clau|AEV95222.1 ldam|AH70_04965 pent|YP_804605
is0655: clau|AEV95223.1 ldam|AH70_04960 pent|YP_804606
is0656: clau|AEV95224.1 ldam|AH70_04955 pent|YP_804607
is0657: clau|AEV95226.1 ldam|AH70_02765 pent|YP_804610
is0658: clau|AEV95227.1 ldam|AH70_02770 pent|YP_804611
is0659: clau|AEV95228.1 ldam|AH70_02780 pent|YP_804612
is0660: clau|AEV95229.1 ldam|AH70_02785 pent|YP_804613
is0661: clau|AEV95230.1 ldam|AH70_02790 pent|YP_804614
is0662: clau|AEV95231.1 ldam|AH70_02795 pent|YP_804615
is0663: clau|AEV95232.1 ldam|AH70_02800 pent|YP_804616
is0664: clau|AEV95234.1 ldam|AH70_02820 pent|YP_804619
is0665: clau|AEV95235.1 ldam|AH70_02825 pent|YP_804620
is0666: clau|AEV95236.1 ldam|AH70_10435 pent|YP_804621
is0667: clau|AEV95237.1 ldam|AH70_10430 pent|YP_804622
is0668: clau|AEV95241.1 ldam|AH70_10420 pent|YP_804398
is0669: clau|AEV95242.1 ldam|AH70_10415 pent|YP_804397
is0670: clau|AEV95243.1 ldam|AH70_10410 pent|YP_804396
is0671: clau|AEV95244.1 ldam|AH70_10405 pent|YP_804395
is0672: clau|AEV95245.1 ldam|AH70_10400 pent|YP_804394
is0673: clau|AEV95246.1 ldam|AH70_10395 pent|YP_804393
is0674: clau|AEV95247.1 ldam|AH70_10390 pent|YP_804392
is0675: clau|AEV95248.1 ldam|AH70_10385 pent|YP_804391
is0676: clau|AEV95249.1 ldam|AH70_10380 pent|YP_804390
is0677: clau|AEV95250.1 ldam|AH70_10375 pent|YP_804389
is0678: clau|AEV95251.1 ldam|AH70_10370 pent|YP_804388
is0679: clau|AEV95252.1 ldam|AH70_10365 pent|YP_804387
is0680: clau|AEV95253.1 ldam|AH70_10360 pent|YP_804386
is0681: clau|AEV95254.1 ldam|AH70_10355 pent|YP_804385
is0682: clau|AEV95255.1 ldam|AH70_10350 pent|YP_804384
is0683: clau|AEV95256.1 ldam|AH70_10345 pent|YP_804383
is0684: clau|AEV95257.1 ldam|AH70_10340 pent|YP_804382
is0685: clau|AEV95258.1 ldam|AH70_10335 pent|YP_804381
is0686: clau|AEV95259.1 ldam|AH70_10330 pent|YP_804380
is0687: clau|AEV95260.1 ldam|AH70_10325 pent|YP_804379
is0688: clau|AEV95261.1 ldam|AH70_10320 pent|YP_804378
is0689: clau|AEV95262.1 ldam|AH70_05660 pent|YP_804377
is0690: clau|AEV95264.1 ldam|AH70_05670 pent|YP_804375
is0691: clau|AEV95266.1 ldam|AH70_05690 pent|YP_804373
is0692: clau|AEV95267.1 ldam|AH70_05695 pent|YP_804372
is0693: clau|AEV95268.1 ldam|AH70_05700 pent|YP_804371
is0694: clau|AEV95269.1 ldam|AH70_05710 pent|YP_804160
is0695: clau|AEV95270.1 ldam|AH70_05715 pent|YP_804370
is0696: clau|AEV95271.1 ldam|AH70_05720 pent|YP_804369
is0697: clau|AEV95273.1 ldam|AH70_08860 pent|YP_804443
is0698: clau|AEV95274.1 ldam|AH70_08865 pent|YP_804442
is0699: clau|AEV95275.1 ldam|AH70_08870 pent|YP_804441
is0700: clau|AEV95276.1 ldam|AH70_08875 pent|YP_804440
is0701: clau|AEV95277.1 ldam|AH70_08880 pent|YP_804439
is0702: clau|AEV95278.1 ldam|AH70_08885 pent|YP_804438
is0703: clau|AEV95279.1 ldam|AH70_08890 pent|YP_804437
is0704: clau|AEV95280.1 ldam|AH70_08895 pent|YP_804436
is0705: clau|AEV95281.1 ldam|AH70_08900 pent|YP_804435
is0706: clau|AEV95282.1 ldam|AH70_08905 pent|YP_804434
is0707: clau|AEV95283.1 ldam|AH70_08910 pent|YP_804433
is0708: clau|AEV95284.1 ldam|AH70_08915 pent|YP_804432
is0709: clau|AEV95285.1 ldam|AH70_08920 pent|YP_804431
is0710: clau|AEV95286.1 ldam|AH70_08925 pent|YP_804430
is0711: clau|AEV95287.1 ldam|AH70_08930 pent|YP_804077
is0712: clau|AEV95288.1 ldam|AH70_08935 pent|YP_804429
is0713: clau|AEV95289.1 ldam|AH70_08940 pent|YP_804428
is0714: clau|AEV95290.1 ldam|AH70_08945 pent|YP_804427
is0715: clau|AEV95291.1 ldam|AH70_08950 pent|YP_804426
is0716: clau|AEV95292.1 ldam|AH70_08955 pent|YP_804425
is0717: clau|AEV95293.1 ldam|AH70_08960 pent|YP_804424
is0718: clau|AEV95294.1 ldam|AH70_02595 pent|YP_804416
is0719: clau|AEV95295.1 ldam|AH70_02600 pent|YP_804415
is0720: clau|AEV95296.1 ldam|AH70_02605 pent|YP_804414
is0721: clau|AEV95297.1 ldam|AH70_03390 pent|YP_804413
is0722: clau|AEV95298.1 ldam|AH70_03400 pent|YP_804352
is0723: clau|AEV95299.1 ldam|AH70_03405 pent|YP_804351
is0724: clau|AEV95300.1 ldam|AH70_03410 pent|YP_804350
is0725: clau|AEV95301.1 ldam|AH70_03420 pent|YP_804349
is0726: clau|AEV95302.1 ldam|AH70_03575 pent|YP_804348
is0727: clau|AEV95303.1 ldam|AH70_03570 pent|YP_804347
is0728: clau|AEV95304.1 ldam|AH70_03565 pent|YP_804346
is0729: clau|AEV95305.1 ldam|AH70_03560 pent|YP_804345

is0730: clau|AEV95306.1 ldam|AH70_03555 pent|YP_804344
is0731: clau|AEV95307.1 ldam|AH70_03550 pent|YP_804343
is0732: clau|AEV95308.1 ldam|AH70_03545 pent|YP_804342
is0733: clau|AEV95309.1 ldam|AH70_03540 pent|YP_804341
is0734: clau|AEV95310.1 ldam|AH70_03535 pent|YP_804340
is0735: clau|AEV95311.1 ldam|AH70_03530 pent|YP_804339
is0736: clau|AEV95312.1 ldam|AH70_03525 pent|YP_804338
is0737: clau|AEV95313.1 ldam|AH70_00250 pent|YP_804337
is0738: clau|AEV95314.1 ldam|AH70_00245 pent|YP_804336
is0739: clau|AEV95315.1 ldam|AH70_00240 pent|YP_804335
is0740: clau|AEV95316.1 ldam|AH70_00235 pent|YP_804334
is0741: clau|AEV95317.1 ldam|AH70_00230 pent|YP_804333
is0742: clau|AEV95318.1 ldam|AH70_00225 pent|YP_804332
is0743: clau|AEV95319.1 ldam|AH70_00220 pent|YP_804331
is0744: clau|AEV95320.1 ldam|AH70_00215 pent|YP_804330
is0745: clau|AEV95321.1 ldam|AH70_00210 pent|YP_804329
is0746: clau|AEV95322.1 ldam|AH70_00205 pent|YP_804328
is0747: clau|AEV95323.1 ldam|AH70_00200 pent|YP_804327
is0748: clau|AEV95324.1 ldam|AH70_00195 pent|YP_804326
is0749: clau|AEV95326.1 ldam|AH70_00165 pent|YP_804324
is0750: clau|AEV95327.1 ldam|AH70_00160 pent|YP_804323
is0751: clau|AEV95328.1 ldam|AH70_00155 pent|YP_804322
is0752: clau|AEV95329.1 ldam|AH70_00150 pent|YP_804321
is0753: clau|AEV95330.1 ldam|AH70_00145 pent|YP_804320
is0754: clau|AEV95331.1 ldam|AH70_00140 pent|YP_804319
is0755: clau|AEV95332.1 ldam|AH70_11165 pent|YP_804318
is0756: clau|AEV95333.1 ldam|AH70_11160 pent|YP_804317
is0757: clau|AEV95334.1 ldam|AH70_11155 pent|YP_804316
is0758: clau|AEV95335.1 ldam|AH70_11150 pent|YP_804315
is0759: clau|AEV95336.1 ldam|AH70_11145 pent|YP_804314
is0760: clau|AEV95337.1 ldam|AH70_11140 pent|YP_804313
is0761: clau|AEV95339.1 ldam|AH70_03285 pent|YP_804048
is0762: clau|AEV95346.1 ldam|AH70_11110 pent|YP_804311
is0763: clau|AEV95353.1 ldam|AH70_08215 pent|YP_804245
is0764: clau|AEV95354.1 ldam|AH70_08210 pent|YP_804244
is0765: clau|AEV95355.1 ldam|AH70_08205 pent|YP_804243
is0766: clau|AEV95356.1 ldam|AH70_08200 pent|YP_804242
is0767: clau|AEV95357.1 ldam|AH70_08195 pent|YP_804241
is0768: clau|AEV95358.1 ldam|AH70_08190 pent|YP_804240
is0769: clau|AEV95359.1 ldam|AH70_08185 pent|YP_804239
is0770: clau|AEV95360.1 ldam|AH70_08180 pent|YP_804238
is0771: clau|AEV95361.1 ldam|AH70_08175 pent|YP_804237
is0772: clau|AEV95363.1 ldam|AH70_08170 pent|YP_804236
is0773: clau|AEV95364.1 ldam|AH70_08165 pent|YP_804235
is0774: clau|AEV95365.1 ldam|AH70_08160 pent|YP_804234
is0775: clau|AEV95366.1 ldam|AH70_08155 pent|YP_804233
is0776: clau|AEV95367.1 ldam|AH70_08150 pent|YP_804232
is0777: clau|AEV95368.1 ldam|AH70_08145 pent|YP_804230
is0778: clau|AEV95369.1 ldam|AH70_08140 pent|YP_804229
is0779: clau|AEV95370.1 ldam|AH70_08135 pent|YP_804228
is0780: clau|AEV95371.1 ldam|AH70_08130 pent|YP_804227
is0781: clau|AEV95372.1 ldam|AH70_08125 pent|YP_804226
is0782: clau|AEV95373.1 ldam|AH70_08120 pent|YP_804225
is0783: clau|AEV95374.1 ldam|AH70_08115 pent|YP_804224
is0784: clau|AEV95376.1 ldam|AH70_05960 pent|YP_804222
is0785: clau|AEV95377.1 ldam|AH70_05955 pent|YP_804221
is0786: clau|AEV95378.1 ldam|AH70_05950 pent|YP_804220
is0787: clau|AEV95379.1 ldam|AH70_05945 pent|YP_804219
is0788: clau|AEV95380.1 ldam|AH70_05940 pent|YP_804218
is0789: clau|AEV95381.1 ldam|AH70_05935 pent|YP_804217
is0790: clau|AEV95382.1 ldam|AH70_05930 pent|YP_804216
is0791: clau|AEV95383.1 ldam|AH70_05925 pent|YP_804215
is0792: clau|AEV95384.1 ldam|AH70_05920 pent|YP_804214
is0793: clau|AEV95385.1 ldam|AH70_05915 pent|YP_804213
is0794: clau|AEV95386.1 ldam|AH70_05910 pent|YP_804212
is0795: clau|AEV95387.1 ldam|AH70_05905 pent|YP_804211
is0796: clau|AEV95388.1 ldam|AH70_05900 pent|YP_804210
is0797: clau|AEV95389.1 ldam|AH70_05895 pent|YP_804209
is0798: clau|AEV95390.1 ldam|AH70_05890 pent|YP_804208
is0799: clau|AEV95391.1 ldam|AH70_05885 pent|YP_804207
is0800: clau|AEV95392.1 ldam|AH70_05880 pent|YP_804206
is0801: clau|AEV95393.1 ldam|AH70_05875 pent|YP_804205
is0802: clau|AEV95395.1 ldam|AH70_05870 pent|YP_804204
is0803: clau|AEV95396.1 ldam|AH70_05865 pent|YP_804203
is0804: clau|AEV95397.1 ldam|AH70_05860 pent|YP_804202
is0805: clau|AEV95398.1 ldam|AH70_05855 pent|YP_804197
is0806: clau|AEV95399.1 ldam|AH70_05850 pent|YP_804196
is0807: clau|AEV95400.1 ldam|AH70_05845 pent|YP_804195
is0808: clau|AEV95401.1 ldam|AH70_05835 pent|YP_804194
is0809: clau|AEV95402.1 ldam|AH70_05830 pent|YP_804193
is0810: clau|AEV95403.1 ldam|AH70_05825 pent|YP_804192
is0811: clau|AEV95404.1 ldam|AH70_02590 pent|YP_804191
is0812: clau|AEV95406.1 ldam|AH70_08730 pent|YP_804190
is0813: clau|AEV95407.1 ldam|AH70_08725 pent|YP_804189
is0814: clau|AEV95408.1 ldam|AH70_08720 pent|YP_804188
is0815: clau|AEV95409.1 ldam|AH70_08715 pent|YP_804187
is0816: clau|AEV95410.1 ldam|AH70_08710 pent|YP_804186
is0817: clau|AEV95411.1 ldam|AH70_08705 pent|YP_804185
is0818: clau|AEV95412.1 ldam|AH70_08700 pent|YP_804184
is0819: clau|AEV95413.1 ldam|AH70_08695 pent|YP_804183
is0820: clau|AEV95414.1 ldam|AH70_08690 pent|YP_804182
is0821: clau|AEV95415.1 ldam|AH70_08685 pent|YP_804181
is0822: clau|AEV95416.1 ldam|AH70_08680 pent|YP_804180
is0823: clau|AEV95417.1 ldam|AH70_08675 pent|YP_804179
is0824: clau|AEV95418.1 ldam|AH70_08670 pent|YP_804178
is0825: clau|AEV95419.1 ldam|AH70_08665 pent|YP_804177
is0826: clau|AEV95420.1 ldam|AH70_08660 pent|YP_804176
is0827: clau|AEV95421.1 ldam|AH70_08655 pent|YP_804175
is0828: clau|AEV95422.1 ldam|AH70_08650 pent|YP_804174
is0829: clau|AEV95423.1 ldam|AH70_08645 pent|YP_804173
is0830: clau|AEV95424.1 ldam|AH70_08630 pent|YP_804172
is0831: clau|AEV95428.1 ldam|AH70_08620 pent|YP_804166
is0832: clau|AEV95429.1 ldam|AH70_08615 pent|YP_804165
is0833: clau|AEV95430.1 ldam|AH70_08610 pent|YP_804164
is0834: clau|AEV95431.1 ldam|AH70_08605 pent|YP_804163
is0835: clau|AEV95438.1 ldam|AH70_08570 pent|YP_804159
is0836: clau|AEV95439.1 ldam|AH70_08565 pent|YP_804158
is0837: clau|AEV95441.1 ldam|AH70_08560 pent|YP_804156
is0838: clau|AEV95442.1 ldam|AH70_08590 pent|YP_804155
is0839: clau|AEV95443.1 ldam|AH70_08550 pent|YP_804154
is0840: clau|AEV95444.1 ldam|AH70_08540 pent|YP_804152
is0841: clau|AEV95445.1 ldam|AH70_08535 pent|YP_804151
is0842: clau|AEV95446.1 ldam|AH70_08525 pent|YP_804150
is0843: clau|AEV95447.1 ldam|AH70_08520 pent|YP_804149
is0844: clau|AEV95448.1 ldam|AH70_08515 pent|YP_804148
is0845: clau|AEV95449.1 ldam|AH70_08510 pent|YP_804147
is0846: clau|AEV95451.1 ldam|AH70_08495 pent|YP_804145
is0847: clau|AEV95453.1 ldam|AH70_09230 pent|YP_804143
is0848: clau|AEV95454.1 ldam|AH70_09235 pent|YP_804142
is0849: clau|AEV95455.1 ldam|AH70_09240 pent|YP_804141
is0850: clau|AEV95457.1 ldam|AH70_09245 pent|YP_804139
is0851: clau|AEV95458.1 ldam|AH70_09250 pent|YP_804138
is0852: clau|AEV95459.1 ldam|AH70_09255 pent|YP_804137
is0853: clau|AEV95461.1 ldam|AH70_05390 pent|YP_804135
is0854: clau|AEV95462.1 ldam|AH70_05395 pent|YP_804134
is0855: clau|AEV95464.1 ldam|AH70_09270 pent|YP_804132
is0856: clau|AEV95465.1 ldam|AH70_11030 pent|YP_804131
is0857: clau|AEV95466.1 ldam|AH70_11025 pent|YP_804130
is0858: clau|AEV95467.1 ldam|AH70_11020 pent|YP_804129
is0859: clau|AEV95469.1 ldam|AH70_11010 pent|YP_804128
is0860: clau|AEV95470.1 ldam|AH70_11005 pent|YP_804127
is0861: clau|AEV95473.1 ldam|AH70_02420 pent|YP_804123
is0862: clau|AEV95474.1 ldam|AH70_02435 pent|YP_804122
is0863: clau|AEV95475.1 ldam|AH70_01650 pent|YP_804121
is0864: clau|AEV95476.1 ldam|AH70_00025 pent|YP_804120
is0865: clau|AEV95477.1 ldam|AH70_00030 pent|YP_804119
is0866: clau|AEV95478.1 ldam|AH70_00035 pent|YP_804118
is0867: clau|AEV95479.1 ldam|AH70_00040 pent|YP_804117
is0868: clau|AEV95480.1 ldam|AH70_00045 pent|YP_804114
is0869: clau|AEV95484.1 ldam|AH70_03845 pent|YP_804105
is0870: clau|AEV95489.1 ldam|AH70_01590 pent|YP_804098
is0871: clau|AEV95490.1 ldam|AH70_01585 pent|YP_804097
is0872: clau|AEV95491.1 ldam|AH70_01580 pent|YP_804096
is0873: clau|AEV95492.1 ldam|AH70_01575 pent|YP_804095
is0874: clau|AEV95494.1 ldam|AH70_00065 pent|YP_804846
is0875: clau|AEV95495.1 ldam|AH70_00070 pent|YP_804847
is0876: clau|AEV95496.1 ldam|AH70_00075 pent|YP_804848
is0877: clau|AEV95497.1 ldam|AH70_00080 pent|YP_804849
is0878: clau|AEV95498.1 ldam|AH70_00085 pent|YP_804850
is0879: clau|AEV95499.1 ldam|AH70_00090 pent|YP_804851
is0880: clau|AEV95500.1 ldam|AH70_00095 pent|YP_804852
is0881: clau|AEV95501.1 ldam|AH70_00100 pent|YP_804853
is0882: clau|AEV95502.1 ldam|AH70_00105 pent|YP_804854
is0883: clau|AEV95503.1 ldam|AH70_00110 pent|YP_804855
is0884: clau|AEV95504.1 ldam|AH70_00115 pent|YP_804856
is0885: clau|AEV95505.1 ldam|AH70_00120 pent|YP_804857
is0886: clau|AEV95506.1 ldam|AH70_06410 pent|YP_804865
is0887: clau|AEV95507.1 ldam|AH70_06405 pent|YP_804866
is0888: clau|AEV95508.1 ldam|AH70_06400 pent|YP_804867
is0889: clau|AEV95509.1 ldam|AH70_06395 pent|YP_804868
is0890: clau|AEV95510.1 ldam|AH70_06390 pent|YP_804869
is0891: clau|AEV95511.1 ldam|AH70_06385 pent|YP_804870
is0892: clau|AEV95512.1 ldam|AH70_06380 pent|YP_804871
is0893: clau|AEV95513.1 ldam|AH70_06375 pent|YP_804872
is0894: clau|AEV95514.1 ldam|AH70_06370 pent|YP_804873
is0895: clau|AEV95515.1 ldam|AH70_06365 pent|YP_804874

is0896: clau|AEV95516.1 ldam|AH70_06355 pent|YP_804876
is0897: clau|AEV95517.1 ldam|AH70_06350 pent|YP_804877
is0898: clau|AEV95518.1 ldam|AH70_06345 pent|YP_804878
is0899: clau|AEV95519.1 ldam|AH70_06340 pent|YP_804879
is0900: clau|AEV95520.1 ldam|AH70_06335 pent|YP_804880
is0901: clau|AEV95521.1 ldam|AH70_06330 pent|YP_804881
is0902: clau|AEV95522.1 ldam|AH70_06325 pent|YP_804882
is0903: clau|AEV95523.1 ldam|AH70_06320 pent|YP_804883
is0904: clau|AEV95524.1 ldam|AH70_06315 pent|YP_804884
is0905: clau|AEV95526.1 ldam|AH70_06310 pent|YP_804886
is0906: clau|AEV95527.1 ldam|AH70_06305 pent|YP_804887
is0907: clau|AEV95528.1 ldam|AH70_06300 pent|YP_804888
is0908: clau|AEV95529.1 ldam|AH70_06295 pent|YP_804889
is0909: clau|AEV95530.1 ldam|AH70_06290 pent|YP_804890
is0910: clau|AEV95531.1 ldam|AH70_06285 pent|YP_804891
is0911: clau|AEV95532.1 ldam|AH70_06280 pent|YP_804892
is0912: clau|AEV95533.1 ldam|AH70_06275 pent|YP_804893
is0913: clau|AEV95534.1 ldam|AH70_06270 pent|YP_804894
is0914: clau|AEV95535.1 ldam|AH70_06265 pent|YP_804895
is0915: clau|AEV95536.1 ldam|AH70_06260 pent|YP_804896
is0916: clau|AEV95537.1 ldam|AH70_06255 pent|YP_804897
is0917: clau|AEV95538.1 ldam|AH70_06250 pent|YP_804898
is0918: clau|AEV95539.1 ldam|AH70_06245 pent|YP_804899
is0919: clau|AEV95541.1 ldam|AH70_06235 pent|YP_804900
is0920: clau|AEV95542.1 ldam|AH70_06230 pent|YP_804901
is0921: clau|AEV95543.1 ldam|AH70_06225 pent|YP_804902
is0922: clau|AEV95545.1 ldam|AH70_06215 pent|YP_804904
is0923: clau|AEV95546.1 ldam|AH70_06210 pent|YP_804905
is0924: clau|AEV95548.1 ldam|AH70_06200 pent|YP_804907
is0925: clau|AEV95549.1 ldam|AH70_06150 pent|YP_804909
is0926: clau|AEV95550.1 ldam|AH70_06145 pent|YP_804910
is0927: clau|AEV95551.1 ldam|AH70_06140 pent|YP_804911
is0928: clau|AEV95554.1 ldam|AH70_06120 pent|YP_804914
is0929: clau|AEV95555.1 ldam|AH70_06110 pent|YP_804916
is0930: clau|AEV95556.1 ldam|AH70_06105 pent|YP_804917
is0931: clau|AEV95557.1 ldam|AH70_06100 pent|YP_804918
is0932: clau|AEV95560.1 ldam|AH70_06070 pent|YP_804919
is0933: clau|AEV95561.1 ldam|AH70_06065 pent|YP_804920
is0934: clau|AEV95562.1 ldam|AH70_06060 pent|YP_804921
is0935: clau|AEV95563.1 ldam|AH70_06050 pent|YP_804922
is0936: clau|AEV95565.1 ldam|AH70_06035 pent|YP_804923
is0937: clau|AEV95568.1 ldam|AH70_06000 pent|YP_804092
is0938: clau|AEV95569.1 ldam|AH70_09355 pent|YP_804310
is0939: clau|AEV95574.1 ldam|AH70_02045 pent|YP_804928
is0940: clau|AEV95575.1 ldam|AH70_04385 pent|YP_804087
is0941: clau|AEV95581.1 ldam|AH70_00850 pent|YP_803950
is0942: clau|AEV95583.1 ldam|AH70_00845 pent|YP_803945
is0943: clau|AEV95587.1 ldam|AH70_01530 pent|YP_803941
is0944: clau|AEV95590.1 ldam|AH70_07700 pent|YP_805086
is0945: clau|AEV95633.1 pent|YP_805168 ldam|AH70_01090
is0946: clau|AEV95649.1 ldam|AH70_03745 pent|YP_803938
is0947: clau|AEV95650.1 ldam|AH70_03740 pent|YP_803937
is0948: clau|AEV95651.1 ldam|AH70_03735 pent|YP_803936
is0949: clau|AEV95652.1 ldam|AH70_03730 pent|YP_803935
is0950: clau|AEV95654.1 ldam|AH70_03710 pent|YP_803933
is0951: clau|AEV95659.1 ldam|AH70_01520 pent|YP_803613
is0952: clau|AEV95660.1 ldam|AH70_04465 pent|YP_803699
is0953: clau|AEV95661.1 ldam|AH70_01205 pent|YP_805238
is0954: clau|AEV95662.1 ldam|AH70_01200 pent|YP_805239
is0955: clau|AEV95663.1 ldam|AH70_01195 pent|YP_805240
is0956: clau|AEV95670.1 ldam|AH70_07035 pent|YP_803930
is0957: clau|AEV95671.1 ldam|AH70_07030 pent|YP_803929
is0958: clau|AEV95672.1 ldam|AH70_07025 pent|YP_803928
is0959: clau|AEV95673.1 ldam|AH70_07020 pent|YP_803927
is0960: clau|AEV95674.1 ldam|AH70_07015 pent|YP_803926
is0961: clau|AEV95676.1 ldam|AH70_07005 pent|YP_803924
is0962: clau|AEV95677.1 ldam|AH70_07000 pent|YP_803923
is0963: clau|AEV95678.1 ldam|AH70_06995 pent|YP_803922
is0964: clau|AEV95680.1 ldam|AH70_02385 pent|YP_803920
is0965: clau|AEV95681.1 ldam|AH70_02380 pent|YP_803918
is0966: clau|AEV95682.1 ldam|AH70_02375 pent|YP_805099
is0967: clau|AEV95683.1 ldam|AH70_02740 pent|YP_803917
is0968: clau|AEV95684.1 ldam|AH70_01960 pent|YP_803916
is0969: clau|AEV95686.1 ldam|AH70_02735 pent|YP_803915
is0970: clau|AEV95687.1 ldam|AH70_02730 pent|YP_803914
is0971: clau|AEV95691.1 ldam|AH70_00625 pent|YP_803913
is0972: clau|AEV95692.1 ldam|AH70_00630 pent|YP_803912
is0973: clau|AEV95693.1 ldam|AH70_00635 pent|YP_803911
is0974: clau|AEV95694.1 ldam|AH70_00640 pent|YP_803910
is0975: clau|AEV95695.1 ldam|AH70_07065 pent|YP_803909
is0976: clau|AEV95696.1 ldam|AH70_00655 pent|YP_803908
is0977: clau|AEV95698.1 ldam|AH70_00665 pent|YP_803906
is0978: clau|AEV95700.1 ldam|AH70_00675 pent|YP_803904
is0979: clau|AEV95701.1 ldam|AH70_00690 pent|YP_803903
is0980: clau|AEV95702.1 ldam|AH70_00695 pent|YP_803902
is0981: clau|AEV95704.1 ldam|AH70_00700 pent|YP_803901
is0982: clau|AEV95705.1 ldam|AH70_00705 pent|YP_803900
is0983: clau|AEV95706.1 ldam|AH70_00710 pent|YP_803899
is0984: clau|AEV95709.1 ldam|AH70_00725 pent|YP_803896
is0985: clau|AEV95710.1 ldam|AH70_00730 pent|YP_803895
is0986: clau|AEV95711.1 ldam|AH70_04165 pent|YP_803756
is0987: clau|AEV95712.1 ldam|AH70_00740 pent|YP_803893
is0988: clau|AEV95716.1 ldam|AH70_01355 pent|YP_803891
is0989: clau|AEV95717.1 ldam|AH70_01360 pent|YP_803890
is0990: clau|AEV95718.1 ldam|AH70_01875 pent|YP_803889
is0991: clau|AEV95720.1 ldam|AH70_01895 pent|YP_803888
is0992: clau|AEV95721.1 ldam|AH70_01900 pent|YP_803887
is0993: clau|AEV95723.1 ldam|AH70_10580 pent|YP_803766
is0994: clau|AEV95724.1 ldam|AH70_07785 pent|YP_803764
is0995: clau|AEV95725.1 ldam|AH70_03330 pent|YP_804170
is0996: clau|AEV95726.1 ldam|AH70_10090 pent|YP_805079
is0997: clau|AEV95728.1 ldam|AH70_10060 pent|YP_805081
is0998: clau|AEV95730.1 ldam|AH70_05980 pent|YP_803754
is0999: clau|AEV95731.1 ldam|AH70_04250 pent|YP_805084
is1000: clau|AEV95732.1 ldam|AH70_07695 pent|YP_805085
is1001: clau|AEV95734.1 ldam|AH70_09010 pent|YP_805096
is1002: clau|AEV95735.1 ldam|AH70_08095 pent|YP_803698
is1003: clau|AEV95736.1 ldam|AH70_01675 pent|YP_803585
is1004: clau|AEV95739.1 ldam|AH70_05355 pent|YP_803688
is1005: clau|AEV95742.1 ldam|AH70_05460 pent|YP_804049
is1006: clau|AEV95750.1 ldam|AH70_04350 pent|YP_805101
is1007: clau|AEV95752.1 ldam|AH70_05555 pent|YP_804941
is1008: clau|AEV95759.1 ldam|AH70_08245 pent|YP_803574
is1009: clau|AEV95761.1 ldam|AH70_08235 pent|YP_803576
is1010: clau|AEV95766.1 ldam|AH70_09750 pent|YP_805139
is1011: clau|AEV95767.1 ldam|AH70_00540 pent|YP_805128
is1012: clau|AEV95770.1 ldam|AH70_04650 pent|YP_804839
is1013: clau|AEV95778.1 ldam|AH70_00020 pent|YP_804089
is1014: clau|AEV95785.1 ldam|AH70_08230 pent|YP_803663
is1015: clau|AEV95786.1 ldam|AH70_06555 pent|YP_803696
is1016: clau|AEV95788.1 ldam|AH70_07825 pent|YP_804405
is1017: clau|AEV95792.1 ldam|AH70_05265 pent|YP_805131
is1018: clau|AEV95804.1 ldam|AH70_06750 pent|YP_803761
is1019: clau|AEV95808.1 ldam|AH70_05705 pent|YP_804453
is1020: clau|AEV95820.1 ldam|AH70_01555 pent|YP_803657
is1021: clau|AEV95826.1 ldam|AH70_03340 pent|YP_804093
is1022: clau|AEV95829.1 ldam|AH70_06540 pent|YP_805145
is1023: clau|AEV95830.1 ldam|AH70_06535 pent|YP_805146
is1024: clau|AEV95833.1 pent|YP_803946 ldam|AH70_05450
is1025: clau|AEV95846.1 ldam|AH70_04930 pent|YP_805142
is1026: clau|AEV95853.1 ldam|AH70_04150 pent|YP_805173
is1027: clau|AEV95854.1 ldam|AH70_04145 pent|YP_805174
is1028: clau|AEV95861.1 ldam|AH70_04120 pent|YP_805176
is1029: clau|AEV95862.1 ldam|AH70_04115 pent|YP_805177
is1030: clau|AEV95863.1 ldam|AH70_04105 pent|YP_805178
is1031: clau|AEV95864.1 ldam|AH70_04100 pent|YP_805179
is1032: clau|AEV95865.1 ldam|AH70_04095 pent|YP_805180
is1033: clau|AEV95866.1 ldam|AH70_10780 pent|YP_805181
is1034: clau|AEV95867.1 ldam|AH70_04080 pent|YP_805182
is1035: clau|AEV95869.1 ldam|AH70_04075 pent|YP_805183
is1036: clau|AEV95870.1 ldam|AH70_04070 pent|YP_805184
is1037: clau|AEV95872.1 ldam|AH70_06800 pent|YP_805072
is1038: clau|AEV95873.1 ldam|AH70_04065 pent|YP_805185
is1039: clau|AEV95874.1 ldam|AH70_04060 pent|YP_805186
is1040: clau|AEV95875.1 ldam|AH70_04055 pent|YP_805187
is1041: clau|AEV95881.1 ldam|AH70_03750 pent|YP_803744
is1042: clau|AEV95883.1 ldam|AH70_04050 pent|YP_804539
is1043: clau|AEV95885.1 ldam|AH70_04040 pent|YP_804541
is1044: clau|AEV95890.1 ldam|AH70_04690 pent|YP_805267
is1045: clau|AEV95891.1 ldam|AH70_04685 pent|YP_805268
is1046: clau|AEV95897.1 ldam|AH70_08450 pent|YP_805189
is1047: clau|AEV95898.1 ldam|AH70_08420 pent|YP_805190
is1048: clau|AEV95899.1 ldam|AH70_08415 pent|YP_805191
is1049: clau|AEV95900.1 ldam|AH70_08345 pent|YP_805193
is1050: clau|AEV95901.1 ldam|AH70_08340 pent|YP_805194
is1051: clau|AEV95902.1 ldam|AH70_08325 pent|YP_805195
is1052: clau|AEV95912.1 ldam|AH70_10220 pent|YP_804464
is1053: clau|AEV95929.1 ldam|AH70_01155 pent|YP_803775
is1054: clau|AEV95930.1 ldam|AH70_01160 pent|YP_803774
is1055: clau|AEV95931.1 ldam|AH70_01150 pent|YP_803773
is1056: clau|AEV95935.1 ldam|AH70_05425 pent|YP_805201
is1057: clau|AEV95936.1 ldam|AH70_01965 pent|YP_805202
is1058: clau|AEV95937.1 ldam|AH70_04790 pent|YP_805204
is1059: clau|AEV95944.1 ldam|AH70_06745 pent|YP_805206
is1060: clau|AEV95945.1 ldam|AH70_06740 pent|YP_805207
is1061: clau|AEV95961.1 ldam|AH70_05320 pent|YP_803772

is1062: clau|AEV95973.1 ldam|AH70_09325 pent|YP_805111
is1063: clau|AEV95974.1 ldam|AH70_05340 pent|YP_803645
is1064: clau|AEV95975.1 ldam|AH70_05335 pent|YP_803644
is1065: clau|AEV95976.1 ldam|AH70_05445 pent|YP_805112
is1066: clau|AEV95982.1 ldam|AH70_00190 pent|YP_805115
is1067: clau|AEV95983.1 ldam|AH70_00185 pent|YP_805116
is1068: clau|AEV95984.1 ldam|AH70_00180 pent|YP_805117
is1069: clau|AEV95985.1 ldam|AH70_00175 pent|YP_805118
is1070: clau|AEV95989.1 ldam|AH70_05465 pent|YP_805119
is1071: clau|AEV95999.1 ldam|AH70_05405 pent|YP_805120
is1072: clau|AEV96004.1 ldam|AH70_09810 pent|YP_805209
is1073: clau|AEV96010.1 ldam|AH70_02475 pent|YP_805213
is1074: clau|AEV96011.1 ldam|AH70_02470 pent|YP_805214
is1075: clau|AEV96012.1 ldam|AH70_05975 pent|YP_804074
is1076: clau|AEV96014.1 ldam|AH70_04495 pent|YP_803659
is1077: clau|AEV96021.1 ldam|AH70_01055 pent|YP_805243
is1078: clau|AEV96033.1 ldam|AH70_05330 pent|YP_805102
is1079: clau|AEV96034.1 ldam|AH70_05325 pent|YP_805103
is1080: clau|AEV96037.1 ldam|AH70_11350 pent|YP_805292
is1081: clau|AEV96038.1 ldam|AH70_11345 pent|YP_805255
is1082: clau|AEV96039.1 ldam|AH70_11340 pent|YP_805256
is1083: clau|AEV96040.1 ldam|AH70_11335 pent|YP_805257
is1084: clau|AEV96041.1 ldam|AH70_11330 pent|YP_805258
is1085: clau|AEV96042.1 ldam|AH70_11325 pent|YP_805259
is1086: clau|AEV96043.1 ldam|AH70_11320 pent|YP_805260
is1087: clau|AEV96044.1 ldam|AH70_01025 pent|YP_803955
is1088: clau|AEV96054.1 ldam|AH70_04400 pent|YP_803595
is1089: clau|AEV96055.1 ldam|AH70_04395 pent|YP_803594
is1090: clau|AEV96057.1 ldam|AH70_06700 pent|YP_803593
is1091: clau|AEV96059.1 ldam|AH70_04380 pent|YP_803591
is1092: clau|AEV96060.1 ldam|AH70_04375 pent|YP_805293
is1093: clau|AEV96061.1 ldam|AH70_04370 pent|YP_805294
is1094: clau|AEV96062.1 ldam|AH70_04365 pent|YP_805295
is1095: clau|AEV96064.1 ldam|AH70_04325 pent|YP_805297
is1096: clau|AEV96065.1 ldam|AH70_04320 pent|YP_805298
is1097: clau|AEV96067.1 ldam|AH70_04315 pent|YP_805299
is1098: clau|AEV96068.1 ldam|AH70_04310 pent|YP_805300
is1099: clau|AEV96070.1 ldam|AH70_04295 pent|YP_805301
is1100: clau|AEV96072.1 ldam|AH70_04280 pent|YP_805302
is1101: clau|AEV96075.1 ldam|AH70_04230 pent|YP_805304
is1102: clau|AEV96076.1 ldam|AH70_04225 pent|YP_805305
is1103: clau|AEV96077.1 ldam|AH70_04215 pent|YP_805306
is1104: clau|AEV96078.1 ldam|AH70_04210 pent|YP_805307
is1105: clau|AEV96085.1 ldam|AH70_00375 pent|YP_804862
is1106: clau|AEV96176.1 ldam|AH70_10915 pent|YP_804516
is1107: clau|AEV96200.1 ldam|AH70_11220 pent|YP_804725
is1108: clau|AEV96201.1 ldam|AH70_08320 pent|YP_804724
is1109: clau|AEV94411.1 clau|AEV95841.1
is1110: clau|AEV94501.1 clau|AEV96001.1
is1111: clau|AEV94572.1 clau|AEV94573.1
is1112: clau|AEV94583.1 clau|AEV94584.1
is1113: clau|AEV94635.1 clau|AEV94902.1
is1114: clau|AEV94702.1 clau|AEV94703.1
is1115: clau|AEV94722.1 clau|AEV94767.1
is1116: clau|AEV94754.1 clau|AEV95595.1
is1117: clau|AEV94788.1 clau|AEV95837.1
is1118: clau|AEV94789.1 clau|AEV95836.1
is1119: clau|AEV94794.1 clau|AFD55029.1
is1120: clau|AEV94838.1 clau|AEV95209.1
is1121: clau|AEV95140.1 clau|AEV95152.1
is1122: clau|AEV95144.1 clau|AEV95148.1
is1123: clau|AEV95570.1 clau|AEV96181.1
is1124: clau|AEV95598.1 clau|AEV95630.1
is1125: clau|AEV95613.1 clau|AEV96168.1
is1126: clau|AEV95614.1 clau|AEV95616.1
is1127: clau|AEV95703.1 clau|AEV95729.1
is1128: clau|AEV95755.1 clau|AEV95942.1
is1129: clau|AEV96084.1 clau|AEV96094.1
is1130: clau|AEV96103.1 clau|AFD55018.1
is1131: clau|AEV96106.1 clau|AEV96161.1
is1132: clau|AEV96107.1 clau|AEV96160.1
is1133: clau|AEV96116.1 clau|AFD55021.1
is1134: clau|AEV96151.1 clau|AEV96166.1
is1135: ldam|AH70_00005 ldam|AH70_01020
is1136: ldam|AH70_00875 ldam|AH70_05645
is1137: ldam|AH70_00895 ldam|AH70_00905
is1138: ldam|AH70_01050 ldam|AH70_03880
is1139: ldam|AH70_01105 ldam|AH70_06805
is1140: ldam|AH70_01310 ldam|AH70_06595
is1141: ldam|AH70_01445 ldam|AH70_01450
is1142: ldam|AH70_02065 ldam|AH70_10305
is1143: ldam|AH70_02085 ldam|AH70_03670
is1144: ldam|AH70_02325 ldam|AH70_05255
is1145: ldam|AH70_02865 ldam|AH70_08110
is1146: ldam|AH70_02870 ldam|AH70_07835
is1147: ldam|AH70_03800 ldam|AH70_03810
is1148: ldam|AH70_04305 ldam|AH70_06840
is1149: ldam|AH70_04410 ldam|AH70_04420
is1150: ldam|AH70_04505 ldam|AH70_04510
is1151: ldam|AH70_05740 ldam|AH70_07090
is1152: ldam|AH70_08385 ldam|AH70_08390
is1153: ldam|AH70_08460 ldam|AH70_08820
is1154: pent|YP_803667 pent|YP_804546
is1155: pent|YP_803705 pent|YP_803731
is1156: pent|YP_803726 pent|YP_803742
is1157: pent|YP_803727 pent|YP_803743
is1158: pent|YP_804046 pent|YP_803746
is1159: pent|YP_804060 pent|YP_805233
is1160: pent|YP_804231 pent|YP_804969
is1161: pent|YP_804250 pent|YP_804533
is1162: pent|YP_804260 pent|YP_804520
is1163: pent|YP_804261 pent|YP_804519
is1164: pent|YP_804266 pent|YP_804513
is1165: pent|YP_804268 pent|YP_804512
is1166: pent|YP_804269 pent|YP_804511
is1167: pent|YP_804270 pent|YP_804508
is1168: pent|YP_804408 pent|YP_804528
is1169: pent|YP_804417 pent|YP_805211
is1170: pent|YP_804828 pent|YP_804829
is1171: pent|YP_805090 pent|YP_805091
is1172: pent|YP_804255 pent|YP_804527
is1173: pent|YP_804257 pent|YP_804525
is1174: pent|YP_804258 pent|YP_804522
is1175: pent|YP_804265 pent|YP_804514
is1176: pent|YP_804406 pent|YP_805251
is1177: pent|YP_804858 pent|YP_805165
is1178: ldam|AH70_01275 ldam|AH70_02150
is1179: clau|AEV95619.1 clau|AEV95623.1
is1180: clau|AEV96171.1 clau|AEV96172.1
is1181: clau|AEV94339.1 pent|YP_805308
is1182: clau|AEV94353.1 ldam|AH70_10695
is1183: clau|AEV94360.1 ldam|AH70_00285
is1184: clau|AEV94376.1 pent|YP_805217
is1185: clau|AEV94388.1 pent|YP_803777
is1186: clau|AEV94396.1 ldam|AH70_09655
is1187: clau|AEV94404.1 pent|YP_803785
is1188: clau|AEV94405.1 pent|YP_803786
is1189: clau|AEV94417.1 pent|YP_803803
is1190: clau|AEV94422.1 pent|YP_803808
is1191: clau|AEV94431.1 ldam|AH70_00880
is1192: clau|AEV94432.1 pent|YP_804355
is1193: clau|AEV94433.1 pent|YP_804356
is1194: clau|AEV94434.1 pent|YP_804357
is1195: clau|AEV94436.1 pent|YP_804359
is1196: clau|AEV94437.1 pent|YP_804360
is1197: clau|AEV94439.1 pent|YP_804362
is1198: clau|AEV94443.1 pent|YP_804366
is1199: clau|AEV94445.1 pent|YP_805100
is1200: clau|AEV94447.1 pent|YP_803608
is1201: clau|AEV94449.1 ldam|AH70_09685
is1202: clau|AEV94456.1 pent|YP_803859
is1203: clau|AEV94457.1 pent|YP_803860
is1204: clau|AEV94458.1 pent|YP_803861
is1205: clau|AEV94459.1 pent|YP_803862
is1206: clau|AEV94460.1 pent|YP_803857
is1207: clau|AEV94461.1 pent|YP_803669
is1208: clau|AEV94468.1 ldam|AH70_08440
is1209: clau|AEV94479.1 pent|YP_803675
is1210: clau|AEV94480.1 pent|YP_803676
is1211: clau|AEV94481.1 pent|YP_803677
is1212: clau|AEV94490.1 ldam|AH70_02660
is1213: clau|AEV94491.1 ldam|AH70_02665
is1214: pent|YP_804076 ldam|AH70_01000
is1215: clau|AEV94492.1 ldam|AH70_10140
is1216: clau|AEV94496.1 ldam|AH70_10500
is1217: clau|AEV94500.1 pent|YP_804161
is1218: clau|AEV94502.1 pent|YP_805241
is1219: clau|AEV94504.1 pent|YP_805149
is1220: clau|AEV94507.1 pent|YP_803678
is1221: clau|AEV94510.1 pent|YP_804124
is1222: clau|AEV94521.1 ldam|AH70_01100
is1223: clau|AEV94530.1 pent|YP_805042
is1224: clau|AEV94533.1 pent|YP_803640
is1225: clau|AEV94535.1 ldam|AH70_05285
is1226: clau|AEV94567.1 ldam|AH70_07365
is1227: clau|AEV94581.1 ldam|AH70_03460

is1228: clau|AEV94587.1 ldam|AH70_02845
is1229: clau|AEV94601.1 pent|YP_805002
is1230: clau|AEV94610.1 pent|YP_803583
is1231: clau|AEV94611.1 pent|YP_803584
is1232: clau|AEV94612.1 ldam|AH70_02620
is1233: clau|AEV94615.1 pent|YP_803581
is1234: clau|AEV94616.1 ldam|AH70_06985
is1235: clau|AEV94619.1 ldam|AH70_04245
is1236: clau|AEV94624.1 ldam|AH70_05535
is1237: clau|AEV94633.1 ldam|AH70_05345
is1238: clau|AEV94644.1 ldam|AH70_04020
is1239: clau|AEV94645.1 ldam|AH70_04025
is1240: clau|AEV94646.1 pent|YP_804982
is1241: clau|AEV94650.1 pent|YP_804985
is1242: clau|AEV94658.1 ldam|AH70_06565
is1243: clau|AEV94661.1 ldam|AH70_10165
is1244: clau|AEV94672.1 ldam|AH70_04910
is1245: clau|AEV94687.1 pent|YP_804162
is1246: clau|AEV94689.1 pent|YP_803877
is1247: clau|AEV94692.1 ldam|AH70_09305
is1248: clau|AEV94694.1 ldam|AH70_01985
is1249: clau|AEV94699.1 ldam|AH70_01975
is1250: clau|AEV94705.1 ldam|AH70_01700
is1251: clau|AEV94706.1 pent|YP_804566
is1252: clau|AEV94707.1 pent|YP_804565
is1253: clau|AEV94713.1 pent|YP_803771
is1254: clau|AEV94716.1 ldam|AH70_07550
is1255: clau|AEV94724.1 ldam|AH70_08270
is1256: clau|AEV94725.1 ldam|AH70_00330
is1257: clau|AEV94726.1 ldam|AH70_00325
is1258: clau|AEV94727.1 ldam|AH70_00320
is1259: clau|AEV94736.1 ldam|AH70_09745
is1260: clau|AEV94737.1 ldam|AH70_09740
is1261: clau|AEV94738.1 ldam|AH70_09725
is1262: clau|AEV94739.1 ldam|AH70_09720
is1263: clau|AEV94740.1 ldam|AH70_09715
is1264: clau|AEV94744.1 ldam|AH70_07530
is1265: clau|AEV94748.1 ldam|AH70_07500
is1266: clau|AEV94749.1 ldam|AH70_07495
is1267: clau|AEV94751.1 pent|YP_803864
is1268: clau|AEV94780.1 ldam|AH70_07165
is1269: clau|AEV94785.1 pent|YP_803855
is1270: clau|AEV94787.1 ldam|AH70_10010
is1271: clau|AEV94792.1 pent|YP_804961
is1272: pent|YP_804057 ldam|AH70_02855
is1273: clau|AEV94796.1 pent|YP_804959
is1274: clau|AEV94802.1 pent|YP_804953
is1275: clau|AEV94856.1 pent|YP_804841
is1276: clau|AEV94857.1 pent|YP_804842
is1277: clau|AEV94858.1 pent|YP_804843
is1278: clau|AEV94859.1 pent|YP_804844
is1279: clau|AEV94879.1 pent|YP_805253
is1280: clau|AEV94880.1 pent|YP_805252
is1281: clau|AEV94881.1 ldam|AH70_06855
is1282: clau|AEV94883.1 pent|YP_804072
is1283: clau|AEV94884.1 pent|YP_804071
is1284: clau|AEV94886.1 pent|YP_804070
is1285: clau|AEV94887.1 ldam|AH70_10270
is1286: clau|AEV94888.1 ldam|AH70_10275
is1287: clau|AEV94893.1 pent|YP_804505
is1288: clau|AEV94900.1 ldam|AH70_05815
is1289: clau|AEV94904.1 ldam|AH70_08355
is1290: clau|AEV94905.1 ldam|AH70_08360
is1291: clau|AEV94906.1 pent|YP_804837
is1292: clau|AEV94909.1 pent|YP_803815
is1293: clau|AEV94914.1 pent|YP_803616
is1294: clau|AEV94915.1 pent|YP_803617
is1295: clau|AEV94916.1 pent|YP_803618
is1296: clau|AEV94917.1 pent|YP_803619
is1297: clau|AEV94918.1 pent|YP_803620
is1298: clau|AEV94919.1 pent|YP_803621
is1299: clau|AEV94931.1 pent|YP_803598
is1300: clau|AEV94932.1 pent|YP_803599
is1301: clau|AEV94935.1 pent|YP_803762
is1302: clau|AEV94936.1 pent|YP_804823
is1303: clau|AEV94962.1 pent|YP_804797
is1304: clau|AEV94968.1 ldam|AH70_07360
is1305: clau|AEV94973.1 pent|YP_804787
is1306: clau|AEV94994.1 pent|YP_804766
is1307: clau|AEV94998.1 pent|YP_804762
is1308: clau|AEV95007.1 pent|YP_804753
is1309: clau|AEV95023.1 pent|YP_804737
is1310: clau|AEV95028.1 pent|YP_804732
is1311: clau|AEV95029.1 pent|YP_804731
is1312: clau|AEV95052.1 ldam|AH70_02360
is1313: clau|AEV95087.1 pent|YP_804674
is1314: clau|AEV95113.1 pent|YP_804649
is1315: clau|AEV95115.1 ldam|AH70_00430
is1316: clau|AEV95155.1 ldam|AH70_09340
is1317: clau|AEV95156.1 pent|YP_804597
is1318: clau|AEV95176.1 ldam|AH70_05110
is1319: clau|AEV95196.1 pent|YP_804461
is1320: clau|AEV95207.1 pent|YP_804448
is1321: clau|AEV95208.1 ldam|AH70_01060
is1322: clau|AEV95210.1 ldam|AH70_05765
is1323: clau|AEV95213.1 pent|YP_804445
is1324: clau|AEV95217.1 pent|YP_804600
is1325: clau|AEV95218.1 pent|YP_804601
is1326: clau|AEV95225.1 pent|YP_804608
is1327: clau|AEV95233.1 pent|YP_804617
is1328: clau|AEV95263.1 pent|YP_804376
is1329: clau|AEV95265.1 pent|YP_804374
is1330: clau|AEV95325.1 pent|YP_804325
is1331: clau|AEV95340.1 pent|YP_803673
is1332: clau|AEV95344.1 pent|YP_804588
is1333: clau|AEV95347.1 ldam|AH70_05560
is1334: clau|AEV95348.1 ldam|AH70_05565
is1335: clau|AEV95349.1 ldam|AH70_05570
is1336: clau|AEV95350.1 ldam|AH70_05575
is1337: clau|AEV95352.1 ldam|AH70_05585
is1338: clau|AEV95375.1 pent|YP_804223
is1339: clau|AEV95394.1 pent|YP_804034
is1340: clau|AEV95405.1 ldam|AH70_00275
is1341: clau|AEV95427.1 ldam|AH70_08555
is1342: clau|AEV95434.1 ldam|AH70_01140
is1343: clau|AEV95450.1 pent|YP_804146
is1344: clau|AEV95472.1 pent|YP_804125
is1345: clau|AEV95493.1 pent|YP_804094
is1346: clau|AEV95525.1 pent|YP_804885
is1347: clau|AEV95544.1 ldam|AH70_06220
is1348: clau|AEV95552.1 pent|YP_804912
is1349: clau|AEV95553.1 pent|YP_804913
is1350: clau|AEV95558.1 ldam|AH70_06095
is1351: clau|AEV95559.1 ldam|AH70_06085
is1352: clau|AEV95564.1 ldam|AH70_06045
is1353: clau|AEV95566.1 pent|YP_804924
is1354: clau|AEV95567.1 pent|YP_804925
is1355: clau|AEV95571.1 pent|YP_804926
is1356: clau|AEV95573.1 ldam|AH70_10955
is1357: clau|AEV95584.1 pent|YP_803944
is1358: clau|AEV95585.1 pent|YP_803943
is1359: clau|AEV95637.1 ldam|AH70_04360
is1360: clau|AEV95667.1 pent|YP_805161
is1361: clau|AEV95669.1 pent|YP_805163
is1362: ldam|AH70_03515 pent|YP_804053
is1363: clau|AEV95679.1 pent|YP_803921
is1364: clau|AEV95685.1 ldam|AH70_05385
is1365: clau|AEV95689.1 ldam|AH70_00560
is1366: clau|AEV95690.1 ldam|AH70_00565
is1367: clau|AEV95697.1 pent|YP_803907
is1368: clau|AEV95707.1 pent|YP_803898
is1369: clau|AEV95713.1 pent|YP_803570
is1370: clau|AEV95714.1 pent|YP_803572
is1371: clau|AEV95715.1 pent|YP_803892
is1372: clau|AEV95719.1 ldam|AH70_10195
is1373: clau|AEV95722.1 ldam|AH70_01905
is1374: clau|AEV95727.1 pent|YP_805080
is1375: clau|AEV95737.1 ldam|AH70_05755
is1376: clau|AEV95741.1 ldam|AH70_01350
is1377: clau|AEV95743.1 pent|YP_804031
is1378: clau|AEV95744.1 ldam|AH70_04340
is1379: clau|AEV95745.1 ldam|AH70_04335
is1380: clau|AEV95746.1 ldam|AH70_04330
is1381: clau|AEV95748.1 ldam|AH70_09805
is1382: clau|AEV95749.1 ldam|AH70_08785
is1383: clau|AEV95751.1 ldam|AH70_06550
is1384: clau|AEV95753.1 ldam|AH70_07815
is1385: clau|AEV95756.1 ldam|AH70_03090
is1386: clau|AEV95764.1 pent|YP_805107
is1387: clau|AEV95768.1 pent|YP_805129
is1388: clau|AEV95769.1 ldam|AH70_09695
is1389: clau|AEV95773.1 pent|YP_804091
is1390: clau|AEV95777.1 pent|YP_804090
is1391: clau|AEV95779.1 pent|YP_804088
is1392: clau|AEV95781.1 pent|YP_805248
is1393: clau|AEV95789.1 pent|YP_805134

is1394: clau|AEV95790.1 pent|YP_805133
is1395: clau|AEV95791.1 pent|YP_805132
is1396: clau|AEV95793.1 pent|YP_805130
is1397: clau|AEV95800.1 ldam|AH70_01765
is1398: clau|AEV95801.1 ldam|AH70_01760
is1399: clau|AEV95803.1 pent|YP_803886
is1400: clau|AEV95805.1 pent|YP_803847
is1401: clau|AEV95806.1 pent|YP_803846
is1402: clau|AEV95809.1 pent|YP_804419
is1403: clau|AEV95810.1 pent|YP_804420
is1404: clau|AEV95811.1 ldam|AH70_05315
is1405: clau|AEV95812.1 ldam|AH70_10735
is1406: clau|AEV95814.1 ldam|AH70_10740
is1407: clau|AEV95815.1 ldam|AH70_10745
is1408: clau|AEV95817.1 pent|YP_805285
is1409: clau|AEV95818.1 pent|YP_805284
is1410: clau|AEV95819.1 ldam|AH70_01545
is1411: clau|AEV95825.1 ldam|AH70_02395
is1412: clau|AEV95834.1 pent|YP_804545
is1413: clau|AEV95838.1 ldam|AH70_08790
is1414: clau|AEV95840.1 pent|YP_805271
is1415: clau|AEV95847.1 pent|YP_805143
is1416: clau|AEV95850.1 ldam|AH70_02390
is1417: clau|AEV95859.1 ldam|AH70_01265
is1418: clau|AEV95876.1 pent|YP_803751
is1419: clau|AEV95877.1 pent|YP_803662
is1420: clau|AEV95880.1 pent|YP_804047
is1421: clau|AEV95882.1 ldam|AH70_05380
is1422: clau|AEV95886.1 ldam|AH70_00890
is1423: clau|AEV95914.1 ldam|AH70_07770
is1424: clau|AEV95917.1 ldam|AH70_07760
is1425: clau|AEV95918.1 ldam|AH70_07755
is1426: clau|AEV95919.1 ldam|AH70_07750
is1427: clau|AEV95920.1 pent|YP_803740
is1428: clau|AEV95921.1 pent|YP_803739
is1429: clau|AEV95922.1 pent|YP_803738
is1430: clau|AEV95923.1 pent|YP_803737
is1431: clau|AEV95924.1 pent|YP_803736
is1432: clau|AEV95925.1 pent|YP_803735
is1433: clau|AEV95926.1 ldam|AH70_08290
is1434: clau|AEV95927.1 ldam|AH70_08285
is1435: clau|AEV95928.1 ldam|AH70_08280
is1436: clau|AEV95933.1 pent|YP_805200
is1437: clau|AEV95940.1 ldam|AH70_01315
is1438: clau|AEV95943.1 pent|YP_805205
is1439: clau|AEV95948.1 pent|YP_805088
is1440: clau|AEV95949.1 pent|YP_805123
is1441: clau|AEV95950.1 pent|YP_805124
is1442: clau|AEV95951.1 pent|YP_805125
is1443: clau|AEV95952.1 ldam|AH70_00950
is1444: clau|AEV95956.1 ldam|AH70_06585
is1445: clau|AEV95957.1 pent|YP_803597
is1446: clau|AEV95960.1 ldam|AH70_03085
is1447: clau|AEV95977.1 ldam|AH70_07070
is1448: clau|AEV95978.1 pent|YP_805113
is1449: ldam|AH70_08770 pent|YP_805274
is1450: clau|AEV95991.1 pent|YP_805152
is1451: clau|AEV95992.1 pent|YP_805153
is1452: clau|AEV95993.1 pent|YP_805154
is1453: clau|AEV95994.1 pent|YP_805155
is1454: clau|AEV95995.1 pent|YP_805156
is1455: clau|AEV95996.1 pent|YP_805157
is1456: clau|AEV95997.1 pent|YP_805158
is1457: clau|AEV95998.1 pent|YP_805159
is1458: clau|AEV96003.1 ldam|AH70_05750
is1459: clau|AEV96008.1 pent|YP_803952
is1460: clau|AEV96009.1 ldam|AH70_04435
is1461: clau|AEV96013.1 ldam|AH70_05770
is1462: clau|AEV96017.1 ldam|AH70_04475
is1463: clau|AEV96019.1 pent|YP_803697
is1464: clau|AEV96023.1 pent|YP_805246
is1465: clau|AEV96027.1 ldam|AH70_01670
is1466: clau|AEV96029.1 pent|YP_805303
is1467: clau|AEV96036.1 pent|YP_804064
is1468: clau|AEV96053.1 ldam|AH70_05375
is1469: clau|AEV96071.1 ldam|AH70_04285
is1470: clau|AEV96089.1 ldam|AH70_02035
is1471: clau|AEV96097.1 ldam|AH70_06845
is1472: clau|AEV96100.1 ldam|AH70_03630
is1473: clau|AEV96101.1 ldam|AH70_03625
is1474: clau|AEV96104.1 ldam|AH70_11075
is1475: clau|AEV96108.1 ldam|AH70_01865
is1476: clau|AEV96111.1 ldam|AH70_10710
is1477: clau|AEV96120.1 ldam|AH70_00010
is1478: clau|AEV96141.1 ldam|AH70_11260
is1479: clau|AEV96150.1 ldam|AH70_03070
is1480: clau|AEV96152.1 ldam|AH70_00585
is1481: clau|AEV96153.1 ldam|AH70_00580
is1482: clau|AEV96177.1 ldam|AH70_10920
is1483: clau|AEV96180.1 ldam|AH70_03150
is1484: clau|AEV96187.1 ldam|AH70_10310
is1485: clau|AEV96191.1 ldam|AH70_09615
is1486: clau|AEV96193.1 ldam|AH70_02080
is1487: clau|AEV96195.1 ldam|AH70_02070
is1488: clau|AEV96196.1 ldam|AH70_11205
is1489: clau|AEV96197.1 ldam|AH70_11210
is1490: clau|AEV96202.1 ldam|AH70_03335
is1491: clau|AEV96204.1 ldam|AH70_11240
is1492: clau|AEV96205.1 ldam|AH70_11245
is1493: ldam|AH70_00130 pent|YP_803672
is1494: ldam|AH70_00290 pent|YP_803610
is1495: ldam|AH70_00295 pent|YP_804808
is1496: ldam|AH70_00300 pent|YP_804807
is1497: ldam|AH70_00315 pent|YP_803668
is1498: ldam|AH70_00350 pent|YP_803825
is1499: ldam|AH70_00380 pent|YP_803701
is1500: ldam|AH70_00385 pent|YP_803700
is1501: ldam|AH70_00550 pent|YP_805127
is1502: ldam|AH70_00595 pent|YP_805279
is1503: ldam|AH70_00685 pent|YP_804554
is1504: ldam|AH70_00920 pent|YP_805242
is1505: ldam|AH70_00960 pent|YP_803614
is1506: ldam|AH70_00965 pent|YP_804368
is1507: ldam|AH70_00975 pent|YP_805231
is1508: ldam|AH70_00990 pent|YP_803650
is1509: ldam|AH70_01340 pent|YP_803631
is1510: ldam|AH70_01390 pent|YP_804099
is1511: ldam|AH70_01425 pent|YP_803723
is1512: ldam|AH70_01430 pent|YP_803722
is1513: ldam|AH70_01435 pent|YP_803721
is1514: ldam|AH70_01440 pent|YP_803720
is1515: ldam|AH70_01500 pent|YP_803629
is1516: ldam|AH70_01565 pent|YP_804056
is1517: ldam|AH70_01715 pent|YP_803586
is1518: ldam|AH70_01725 pent|YP_804943
is1519: ldam|AH70_01930 pent|YP_805106
is1520: ldam|AH70_02405 pent|YP_803947
is1521: ldam|AH70_02410 pent|YP_803949
is1522: ldam|AH70_02640 pent|YP_805003
is1523: ldam|AH70_02680 pent|YP_803580
is1524: ldam|AH70_02760 pent|YP_804609
is1525: ldam|AH70_02815 pent|YP_804618
is1526: ldam|AH70_02875 pent|YP_804467
is1527: ldam|AH70_03220 pent|YP_804827
is1528: ldam|AH70_03510 pent|YP_804052
is1529: ldam|AH70_03780 pent|YP_805219
is1530: ldam|AH70_04290 pent|YP_803866
is1531: ldam|AH70_04455 pent|YP_805083
is1532: ldam|AH70_05295 pent|YP_803611
is1533: ldam|AH70_05455 pent|YP_803948
is1534: ldam|AH70_05495 pent|YP_804478
is1535: ldam|AH70_05500 pent|YP_804477
is1536: ldam|AH70_05745 pent|YP_803814
is1537: ldam|AH70_06360 pent|YP_804875
is1538: ldam|AH70_06420 pent|YP_803872
is1539: ldam|AH70_06425 pent|YP_803871
is1540: ldam|AH70_06430 pent|YP_803870
is1541: ldam|AH70_06435 pent|YP_803869
is1542: ldam|AH70_06465 pent|YP_804354
is1543: ldam|AH70_06685 pent|YP_804782
is1544: ldam|AH70_06850 pent|YP_805126
is1545: ldam|AH70_07120 pent|YP_805196
is1546: ldam|AH70_07125 pent|YP_805197
is1547: ldam|AH70_07130 pent|YP_805198
is1548: ldam|AH70_07135 pent|YP_805199
is1549: ldam|AH70_07190 pent|YP_803856
is1550: ldam|AH70_07195 pent|YP_805261
is1551: ldam|AH70_07200 pent|YP_805263
is1552: ldam|AH70_07295 pent|YP_804029
is1553: ldam|AH70_07475 pent|YP_804018
is1554: ldam|AH70_07810 pent|YP_805040
is1555: ldam|AH70_07860 pent|YP_805038
is1556: ldam|AH70_08050 pent|YP_804112
is1557: ldam|AH70_08090 pent|YP_803592
is1558: ldam|AH70_08105 pent|YP_804845
is1559: ldam|AH70_08225 pent|YP_804059

is1560: ldam|AH70_08305 pent|YP_803660
is1561: ldam|AH70_08335 pent|YP_803894
is1562: ldam|AH70_08365 pent|YP_804168
is1563: ldam|AH70_08410 pent|YP_805192
is1564: ldam|AH70_08545 pent|YP_804153
is1565: ldam|AH70_08775 pent|YP_805275
is1566: ldam|AH70_09005 pent|YP_804701
is1567: ldam|AH70_09060 pent|YP_805247
is1568: ldam|AH70_09345 pent|YP_804481
is1569: ldam|AH70_09350 pent|YP_804482
is1570: ldam|AH70_09360 pent|YP_804484
is1571: ldam|AH70_09380 pent|YP_804489
is1572: ldam|AH70_09385 pent|YP_804490
is1573: ldam|AH70_09400 pent|YP_804492
is1574: ldam|AH70_09405 pent|YP_804493
is1575: ldam|AH70_09410 pent|YP_804494
is1576: ldam|AH70_09415 pent|YP_804495
is1577: ldam|AH70_09420 pent|YP_804496
is1578: ldam|AH70_09425 pent|YP_804497
is1579: ldam|AH70_09440 pent|YP_804500
is1580: ldam|AH70_09450 pent|YP_804502
is1581: ldam|AH70_09560 pent|YP_804252
is1582: ldam|AH70_09760 pent|YP_805140
is1583: ldam|AH70_09765 pent|YP_803741
is1584: ldam|AH70_09795 pent|YP_803730
is1585: ldam|AH70_09930 pent|YP_803978
is1586: ldam|AH70_09965 pent|YP_803985
is1587: ldam|AH70_10085 pent|YP_803752
is1588: ldam|AH70_10225 pent|YP_803745
is1589: ldam|AH70_10235 pent|YP_804466
is1590: ldam|AH70_10480 pent|YP_804994
is1591: ldam|AH70_10890 pent|YP_805054
is1592: ldam|AH70_11045 pent|YP_803788
is1593: ldam|AH70_11050 pent|YP_803789

Part IV

General Discussion and Future Perspectives

Preamble

The goals of this thesis were to evaluate the state-of-the-art approaches for microbial diversity analyses of food ecosystems and for the description of novel LAB taxa. This Part presents a GENERAL DISCUSSION of the results obtained and provides FUTURE PERSPECTIVES.

In **Chapter 6**, the effectiveness of culture-dependent and culture-independent molecular approaches, traditionally used in microbial diversity analyses of food ecosystems, was assessed. Furthermore, the value of HT sequencing technologies, which were applied to unravel the microbial diversity of mature Belgian red-brown acidic ales (Section 5.1) and to sequence the genome of the dominant bacterial member (*i.e.*, *Pediococcus damnosus* LMG 28219) of this beer ecosystem (Section 5.2), was evaluated. Finally, perspectives on the importance of cultivation and the application of an integrative ecosystem biology approach in future microbial diversity analyses of food ecosystems were provided.

In **Chapter 7**, the value of the traditionally used polyphasic taxonomy approach, which was implemented to characterize several novel LAB species (Sections 4.1, 4.2, and 4.3), was assessed. For the genus *Carnobacterium*, this characterization included the extension of the existing MLSA scheme (Section 4.3). Furthermore, Chapter 7 includes an evaluation of the possibilities of incorporating whole-genome sequences into the description of novel LAB species.

Finally, **Chapter 8** comprises guidelines on how HT sequencing technologies should be integrated in future microbial diversity analyses of food ecosystems and outlines which role the genomic taxonomy could have in these analyses.

6

Considerations for Microbial Diversity Analyses

6.1 Traditional Methods

A variety of culture-dependent and culture-independent molecular techniques are traditionally used in microbial diversity analyses to study the phylogenetic and functional diversity of food ecosystems. The culture-dependent approach, involving plating on a range of empirically chosen media and conditions followed by dereplication using MALDI-TOF MS and identification of representative isolates using 16S rRNA gene sequencing, is labor-intensive and reveals only a fraction of the actual diversity. This is due to the challenging recovery of microorganisms, which are either stressed or entered into a VBNC state (Fleet, 1999; Rappé & Giovannoni, 2003). As a consequence, this empirical approach is inadequate for the complete characterization of microbial communities. On the other hand, culture-independent molecular techniques (*i.e.*, DGGE, RFLP, ARISA, qPCR, FISH, clone libraries, and others) do not depend on the cultivability of microorganisms in a community, but the limitations associated with these methods are their labor-intensity, their low-throughput, and their time-consuming nature. Furthermore, the Sanger DNA-sequencing technology is inadequate for processing complex and/or multiple samples. Analysis of these samples requires the capacity to read DNA from

multiple templates in parallel, which HT sequencing technologies do effectively at low costs.

6.2 Methods Involving HT Sequencing

HT sequencing technologies have become indispensable tools for studying the diversity and the metabolic potential of environmental microorganisms. Because of the biases introduced during microbial diversity analyses applying HT sequencing technologies (Ercolini, 2013; Temperton & Giovannoni, 2012), it is crucial to consider those before launching an experiment and while handling the data. In this chapter, several pre- and post-sequencing considerations are discussed and related to the experimental work done in the framework of the present study.

6.2.1 Pre-Sequencing Considerations

General Biases

As is the case for most of the cultivation-independent methods, it is of utmost importance to obtain metagenomic DNA that represents the actual diversity of a sample of interest. This implies minimizing the number of contaminants and inhibitory substances, while maximizing DNA yield of all community members. The most challenging aspect in the DNA extraction process is cell lysis, as not all members of the community (*i.e.*, Gram-stain-positive bacteria, Gram-stain-negative bacteria, fungi, and/or others) are lysed with the same efficacy, which will inevitably introduce biases in the community composition observed. The DNA extraction protocol, as utilized in Section 5.1, applied a combination of freezing and thawing, enzymatic lysis, and physical breaking to release genomic DNA from the cells and was capable of capturing both Gram-stain-positive (*e.g.*, *Pediococcus*, *Lactobacillus*), Gram-stain-negative (*e.g.*, AAB), and fungal (*e.g.*, *Dekkera*) DNA. An alternative approach is to separate prokaryotic and eukaryotic cells by filtration, followed by the application of DNA extraction procedures optimized for the filtered fractions. However, this alternative approach will distort the actual community structure of the sample. As a consequence of the concerns highlighted above, prior knowledge on the microbial community composition is crucial to determine the optimal DNA extraction procedure for the sample of interest.

Next to the biases associated with metagenomic DNA extraction, applying PCR will inevitably introduce biases to estimations of the actual microbial community diversity. Issues regarding primer universality, chimera formation, PCR errors, non-target and preferential amplification, primer and target-region choice, and others, are common to all PCR-based molecular methods. These biases can be minimized by, for instance, optimization of the PCR protocol, which reduces the incidence of chimera formation by applying a limited number of amplification cycli and a long elongation time (as was applied in Section 5.1). Furthermore, a PCR polymerase with proofreading activity was used in Section 5.1 to minimize the PCR error rate. Another major issue is preferential amplification of shorter fragments when sequencing the ITS region, which shows considerable length variation across fungal taxa. Moreover, the primer and target-region choice will result in phylogenetic analyses into varying taxonomic depths, depending on the length and region of the retrieved partial 16S rRNA gene sequence (Kunin *et al.*, 2008; Engelbrektson *et al.*, 2010).

Habitat

The biodiversity of a habitat, including both species richness (*i.e.*, total number of species) and species evenness (*i.e.*, relative abundance of each of the community members), influences the degree of difficulty of the microbial diversity analyses applying HT sequencing technologies. For metagenomic shotgun sequencing strategies involving assembly, habitats with few abundant microbial species are better targets than habitats with many species of even abundance (Kunin *et al.*, 2008). Furthermore, samples containing a large amount of eukaryotic DNA, as was the case for the Belgian red-brown acidic ale ecosystem (Section 5.1), will require extensive sequencing, elevating the actual costs and complicating downstream data analysis and storage. Because of the issues highlighted above, pre-discovery of the ecosystem of interest is recommended prior to launching metagenomic shotgun sequencing experiments. Preceding microbial diversity analysis using target enrichment strategies in conjunction with abundance estimations (*e.g.*, qPCR) can aid to properly assess the most suited approach to analyze the ecosystem of interest.

As discussed in Section 5.1, the mature Belgian red-brown acidic ale ecosystem was analyzed using 454 pyrosequencing of the partial 16S rRNA gene and ITS1 region and revealed that this ecosystem contained only one dominant bacterial and one dominant fungal member, next to several less abundant microbial members. Subsequent targeted isolation and whole-genome sequencing of the dominant community

members allows the study of the metabolism and gene repertoire of these isolates, providing valuable information on their eco-physiological roles (as was done for the dominant bacterial community member in Section 5.2). In case of cultivation problems with the community members of interest, sequencing the metagenome or the genome of single microbial cells (Kalisky & Quake, 2011) could be valuable alternatives. However, if the major objective of the study is to target the whole community diversity and function, instead of focusing on the dominant members, deep metagenome sequencing may be the preferred solution (van Hijum *et al.*, 2013).

Platform

In earlier times, sequencing of 16S rRNA gene amplicons mostly involved 454 pyrosequencing (as was applied in Section 5.1), whereas now many studies use Illumina sequencing because of its higher throughput and resulting reduced expenses (Caporaso *et al.*, 2010). A valid concern regarding the use of the Illumina platform for sequencing 16S rRNA gene amplicons is that the single-end Illumina reads are less than half the length of the 454 pyrosequencing reads, at least at the time of writing. Nevertheless, Illumina sequencing can take advantage of its higher throughput capacity and its ability to generate paired-end (PE) reads in downstream analyses, enabling PE read assembly into a consensus sequence. Because of the conserved nature of the 16S rRNA gene (1500 bp) (as shown in Sections 4.1, 4.2, and 4.3), sequencing of the partial 16S rRNA gene using 454 pyrosequencing or Illumina sequencing does not allow to distinguish between closely related LAB species and only provides a tentative identification. Therefore, complete or nearly complete sequencing of the 16S rRNA gene would be valuable for future research, with the PacBio RS platform being promising in this respect.

A similar evolution in the choice of platform was apparent for shotgun sequencing of (meta)genomic DNA. The initially most often used 454 pyrosequencing is gradually replaced by the increasing use of Illumina sequencing (as was applied for genome sequencing in Section 5.2), providing shorter reads but a much higher throughput and hence coverage at the same price. A recent comparative study of a freshwater lake planktonic community has shown that Illumina sequencing and 454 pyrosequencing lead to similar results with respect to assemblies and the taxonomic and functional repertoires covered (Luo *et al.*, 2012). Next to 454 pyrosequencing and Illumina sequencing platforms, the PacBio RS platform could be promising, because it generates read lengths that are long enough to span multiple prokaryotic genes

and thus are able to provide reliable genetic contexts and to improve annotations. Despite the high error rate inherent to use of the PacBio RS platform, the errors are mostly random and can be reduced in an almost linear fashion by increased coverage, which contrasts with the inherent systematic errors (*i.e.*, error accumulation at the end of the sequence) of other platforms. Currently, a combination of long- and short-read technologies constitutes a particularly promising approach in future (meta)genomics that bears the potential to significantly advance the field.

Replication

Although HT sequencing technologies bring extensive sample replication and subsequent robust statistical analyses into reach, this is rarely done in microbial ecology studies (Prosser, 2010). It is a good scientific practice to analyze replicates (*i.e.*, biological, technical, and experimental replicates) of a sample and assess whether observed differences are statistically meaningful. Replication implies that researchers should critically assess the techniques and experimental design used and determine whether they are capable of achieving the aims of the experiment (Knight *et al.*, 2012). Nevertheless, sampling replicates is almost impossible because of spatial variability occurring in many habitats (Prosser, 2010). Hence, comparing such alleged replicates may reveal little information on methodological reproducibility.

The research presented in Section 5.1 included biological replicates (*i.e.*, subsequent brews) and technical variation was taken into consideration by performing DNA extractions and PCRs in triplicate and pooling them afterwards. A drawback associated with pooling of replicates is that it destroys spatial variability and makes it impossible to calculate true variability, which is required for comparisons (Prosser, 2010). In Section 5.1, experimental replication was applied for the generous donor sample (*i.e.*, a DNA sample that was previously sequenced) and showed a high reproducibility of 454 pyrosequencing. Similarly, Luo and colleagues (2012) assessed the reproducibility of 454 pyrosequencing and Illumina sequencing by re-analyzing the same microbial community DNA sample and showed that library preparation and sequencing are highly reproducible. Nevertheless, Schloss and coworkers (2011a) measured intra- and inter-sequencing center variation and revealed that their mock community (*i.e.*, artificial community) samples clustered by sequencing center and by sequencing run, demonstrating the value of replicating samples in microbial ecology studies.

Contextual Data

Contextual data are the data associated with microbial diversity analyses, such as habitat description, measures of environmental parameters, sampling procedure, sample storage, and others. The genomic standards consortium (GSC) (Field *et al.*, 2011) has published standards for the minimum information about a (meta)genome sequence [MIMS/MIGS (Field *et al.*, 2008; Kottmann *et al.*, 2008)] and about a marker gene sequence [MIMARKS (Yilmaz *et al.*, 2011)], as part of the minimum information about any sequence (MIxS) standards and checklist (Yilmaz *et al.*, 2011), which are supported by the international nucleotide sequence databases collaboration (INSDC). It is of utmost importance that contextual data are collected and integrated into databases (as was done for the data generated in Sections 5.1 and 5.2), because in the long run these data will allow to extract correlations between geography, time, prevailing environmental conditions, and functions from data that otherwise would never be uncovered (Yilmaz *et al.*, 2011; Delmont *et al.*, 2011; Knight *et al.*, 2012).

6.2.2 Post-Sequencing Considerations

Until recently, sequencing capacity has been the limiting factor in microbial diversity analyses involving HT sequencing. However, the on-going increase in sequencing capacity and resulting reduction of prices have made post-sequen-cing data analysis the main bottleneck. Although progress in sequencing technologies still continues at an exponential pace, data analysis cannot keep up. As a consequence, the cost of sequencing declines continuously, whereas the cost for bioinformatics data analysis increases (Sboner *et al.*, 2011).

Use of the 16S rRNA Gene for Taxonomic Purposes

About one in thousand genes in a metagenomic shotgun sequencing dataset is a 16S rRNA gene and only a fraction of these reads has sufficient length and quality for subsequent phylogenetic analysis. Furthermore, depending on the length and region of the partial 16S rRNA gene sequence retrieved, phylogenetic analysis results into OTUs with varying taxonomic depths (Kunin *et al.*, 2008). Extracting 16S rRNA gene reads from a metagenome dataset is superior to targeted sequencing of 16S rRNA gene amplicons in terms of PCR biases, because no primer selection and amplification are involved for library preparation. Nevertheless,

when used for phylogenetic analysis, metagenomic shotgun sequencing requires extensive sequencing depth, resulting in complex data handling and higher prices compared to target enrichment strategies. Furthermore, the highly conserved nature of the 16S rRNA gene in closely related LAB species complicates identification at species levels (as discussed above). Alternatively, as sequence analysis of protein-encoding genes has proven to be superior for the accurate species level identification of a variety of LAB and AAB [(Naser *et al.*, 2005a; Cleenwerck *et al.*, 2010; De Bruyne *et al.*, 2007) and as shown in Sections 4.1, 4.2, and 4.3], mining a metagenomic shotgun dataset for housekeeping genes could aid in obtaining species-level identifications.

Error Removal

In genome sequencing projects, highly redundant consensus assemblies compensate for sequencing errors (as was applied in Section 5.2). Similarly, assembly of sequences from metagenomic libraries can result in a good draft or even complete genomes (Tyson *et al.*, 2004) when the target species shows little intraspecies variation, but usually requires a substantial amount of sequencing. In addition, assembly of large metagenomic datasets presents challenges with respect to memory requirements. Although the assembly of metagenomes yields longer sequences, it also bears the risk of creating chimeric contigs, particularly in habitats with closely related species or highly conserved sequences occurring across species (Teeling & Glöckner, 2012). Furthermore, assembly distorts abundance information, as overlapping sequences from an abundant species will be identified as belonging to the same genome and consequently joined. This leads to a relative underrepresentation of sequences of abundant species. Hence, gene frequencies are better compared based on read representation rather than on assemblages (Teeling & Glöckner, 2012).

In contrast, target enrichment strategies based on PCR (as applied in Section 5.1) cannot take advantage of consensus assemblies to mask incorrect base calls (Huse *et al.*, 2010). As a consequence, caution must be taken when analyzing datasets that do not rely on assembled reads, because sequencing errors can lead to the artificial inflation of diversity estimates (Kunin *et al.*, 2010) and to erroneous taxonomic assignments. A feasible but imperfect alternative to building consensus sequences for target enrichment strategies is to identify and remove reads that are likely to be incorrect. Different HT sequencing platforms are prone to different types of errors, with homopolymers being frequently associated with 454 pyrosequencing (as was applied in Section 5.1).

Bioinformatics' solutions exist to remove PCR and sequencing errors in target enrichment datasets, with the latter having a higher incidence compared to the former. Schloss and coworkers (2011a) identified several features that are linked to low-quality 454 pyrosequences. For instance, sequences containing homopolymers, ambiguous bases, mismatches in their barcode or primer sequence, and short sequences are more likely to be of low quality. The parameters applied in Section 5.1 were largely similar to those proposed by Schloss and coworkers (2011a), with the exception of several parameters for which more stringent values based on the analysis of a mock community were preferred. Moreover, error accumulation at the end of the reads makes trimming of raw 454 pyrosequencing reads crucial to reduce the overall error rate, although there is still much debate on how stringent this trimming should be (Schloss *et al.*, 2011a; Quince *et al.*, 2011). For instance, the raw reads of approximately 500 bp provided by the 454 GS FLX Titanium system of Roche Diagnostics (as applied in Section 5.1) were trimmed to a length of approximately 250 bp, using the parameter settings proposed by Schloss and coworkers (2011a). Further reduction of the error rate can be achieved by single-linkage pre-clustering (Huse *et al.*, 2010), SeqNoise (Quince *et al.*, 2011) (applied in Section 5.1), or modifications thereof. The analysis of reads generated by the Illumina MiSeq (250 bp) differs in several aspects from those generated by the 454 GS FLX system, because Illumina can take advantage of the PE read assembly, which can be used as a tool to reduce the sequencing error rate. However, Kozich and coworkers (2013) showed that these PE reads should fully overlap to obtain good-quality reads.

Subsequently, the de-noised and trimmed sequences are preferably aligned against a curated reference alignment, which is available for 16S rRNA gene sequences (*e.g.*, the SILVA reference alignment was used in Section 5.1) (Schloss, 2009). Sequences generated due to aspecific amplification and therefore aligning to a wrong region or sequences that were classified as chloroplasts or mitochondria, were removed from the dataset. Furthermore, singletons (*i.e.*, OTUs containing only one sequence) are often sequencing errors and may lead to the artificial inflation of the 'rare biosphere' diversity (*i.e.*, microorganisms present in extremely low abundances) and thereby affect the conclusions (Dickie, 2010; Reeder & Knight, 2009). A lot of debate exists on whether the members of this 'rare biosphere' have ecological significance or should be removed from the dataset (Dickie, 2010; Reeder & Knight, 2009).

Although several steps can be taken prior to sequencing to reduce the rate of chimerism (Acinas *et al.*, 2005; Haas *et al.*, 2011; Thompson *et al.*, 2002) as discussed above, several bioinformatics approaches have been developed to identify and remove chimeras. For instance, Haas and coworkers (2011) developed the ChimeraSlayer algorithm, which was superior to Bellerephon (Huber *et al.*, 2004) and Pintail (Ashelford *et al.*, 2005), relying on a reference database to distinguish chimeric and non-chimeric sequences. Additionally, Quince and colleagues (2011) developed Perseus, which does not use a reference database but requires a training set of sequences similar to the sequences being characterized. Finally, Edgar and coworkers (2011) developed Uchime (as applied in Section 5.1), showing improved performance over ChimeraSlayer, but comparable to that of Perseus.

The mothur software package (`www.mothur.org`, as applied in Section 5.1) is an open-source, platform-independent, community-supported software for describing and comparing microbial communities (Schloss, 2009). It builds upon previous tools [*e.g.*, Amplicon Noise (Quince *et al.*, 2009), single-linkage pre-clustering (Huse *et al.*, 2010), Perseus (Quince *et al.*, 2011), Uchime (Edgar *et al.*, 2011), and others] and provides a flexible and powerful software package for analyzing sequencing data. Furthermore, they present a pipeline for handling 454 and Illumina 16S rRNA gene sequencing data (Schloss *et al.*, 2011a; Kozich *et al.*, 2013). QIIME (Caporaso *et al.*, 2010) (*i.e.*, quantitative insights into microbial ecology) is a valid alternative for the analysis of target enrichment sequencing data.

Database Dependencies

A major concern in microbial diversity analyses involving HT sequencing is that often a significant proportion of the data cannot be assigned to a taxon due to a lack of close matches in reference databases (Qin *et al.*, 2010; Temperton & Giovannoni, 2012). The primary source of annotated nucleotide sequences is the INSDC resource (Nakamura *et al.*, 2013), which includes nucleic acid sequencing data and associated information from the DDBJ (Sugawara *et al.*, 2007), the European nucleotide archive (ENA) (Leinonen *et al.*, 2011), and GenBank (Benson *et al.*, 2013), including the SRA (Kodama *et al.*, 2012) repository. These sequence data collections constitute huge resources of sequence information but, at the same time, their contents are often compromised by identification errors, low-quality sequence data, redundancy, and incompleteness (Turenne *et al.*, 2001). Current functional assignment of genes from (meta)genomes is based on time-consuming homology searches, using BLAST tools (Altschul *et al.*, 1997) (as was applied in

Section 5.2) that depend heavily on the quality and completeness of these available databases. Furthermore, homology-based algorithms are not sensitive, particularly for genomes that lack sequenced relatives and can miss novel genes that may ecologically be the most interesting (Suenaga, 2012). A review of prokaryotic protein diversity in different shotgun metagenomic studies indicated that 30-60% of the proteins cannot be assigned to known functions using current public databases (Vieites *et al.*, 2009), indicating that there is an urgent need to develop specific and quality controlled reference databases that are able to reduce redundancies, validate sequence annotations, and increase accuracy and effectiveness of taxon assignments (Santamaria *et al.*, 2012; Walker *et al.*, 2014).

The approach applied in Section 5.1 for the annotation of the 16S rRNA gene and ITS region sequence data, was based on what has been proposed by Schloss and coworkers (2009; 2010; 2011a). Firstly, reads were binned (*i.e.*, grouping of reads into clusters) into OTUs and subsequently a consensus classification was obtained for each OTU using a trainset (*i.e.*, quality controlled reference sequences), which is available for 16S rRNA gene sequences on the mothur website to train the Naïve Bayesian Classifier (Wang *et al.*, 2007). Unfortunately, no trainset was available for fungal ITS sequences and was therefore developed in-house using reference sequences extracted from curated ITS databases (Weiss *et al.*, 2013; Robert *et al.*, 2013) (as applied in Section 5.1). Another issue complicating the taxonomic assignment of fungal ITS sequences is the lack of means to refer to fungal species, for which no Latin name is available in a standardized way. These shortcomings negatively affect the identification of fungal species based on ITS markers and result in many unclassified fungi, which was apparent in the results of Section 5.1. As a consequence, the development of curated fungal ITS databases will be crucial for future research on fungal diversity.

6.3 Importance of Cultivation

Although purely sequence-based descriptions give valuable information about community content and metabolic potential, they suffer from several inherent limitations (as discussed in Section 6.2). As a consequence, the cultivation of microorganisms will remain essential for future research (Allen-Vercoe, 2013). One benefit of having strains in pure culture is that their genome can be sequenced rapidly and at low-cost per base, thus adding information to reference databases (Walker *et al.*, 2014) and contributing to the interpretation of physiological as well as (meta)genomic data. Furthermore, reference genome data derived from food isolates

can be mined for interesting functionalities to apply in the food industry. Of course, a microorganism can only be applied in the food industry after it has first been grown, isolated, and tested for safety in the laboratory. In addition, it is becoming increasingly clear that microbial ecosystems contain rare phylotypes for which abundance is less important than the ability to serve a key metabolic function upon which the larger community structure may rely (Ze *et al.*, 2013). Culture-based approaches offer the route through which sampling and detailed analysis of these low-abundance species may be accomplished.

Cultivation of microbial community members *in vitro* can be approached in several ways, which are summarized in Figure 22. The first of these is an empirical approach (as discussed in Section 1.1), which traditionally involves the cultivation of microorganisms onto a set of empirically chosen growth media and conditions. The diversity yield using this approach is low because of the enormous range of microbial growth rates and conditions that unknown targets may use, which makes the approach time-consuming and inefficient (Allen-Vercoe, 2013).

Another approach to cultivate microbial communities requires previous knowledge on the ecosystem of interest and involves the use of custom-made growth media for targeted growth of microorganisms that have thus far evaded cultivation (Allen-Vercoe, 2013). In this targeted bioinformatics approach, the molecular signatures of a microbial target are studied before culture is attempted to infer culture conditions. The resulting diversity yield is intentionally low and the target microbiota is efficiently recovered. To our knowledge, this approach has not been applied to a food fermentation ecosystem. An example from research on the gut microbiota is the isolation of *Akkermansia muciniphila* from feces through its unusual ability to use mucin as the sole carbon and nitrogen source (Derrien *et al.*, 2004). An even more targeted approach would be to utilize the genome of a given uncultured species to infer its necessary growth requirements [example publication: (Livermore *et al.*, 2014)].

The third approach to culture microorganisms that thus far evaded cultivation has been developed by Lagier and coworkers (2012) and named culturomics. They started with eliminating the predominant microbial population from stool samples using cocktails of lysogenic phages or filtration methods. This was followed by growth enrichment steps with blood, rumen fluid, or fecal water, and finally 212 separate culture conditions were applied to the resulting pool of microorganisms that had been diluted to give single cells. This approach enabled cultivation of a significant number of low abundant species, which were previously missed by

HT sequencing approaches. The main bottleneck of the culturomics method is the identification step, which is normally performed through time-consuming 16S rRNA gene sequencing of isolates. However, Lagier and colleagues (2012) were able to demonstrate that MALDI-TOF MS (as was applied for dereplication and identification in Sections 4.2 and 4.1, respectively) is a fast and cheap alternative, greatly facilitating the work. Combined with automated colony picking technology and HT growth arrays, culturomics represents a powerful technique that can revolutionize microbial cultivation technology. This approach has the potential to generate a huge amount of bacterial isolates, which may be used to define interactions between bacterial species and between bacterial and eukaryotic cells that are difficult to predict unless tested biologically.

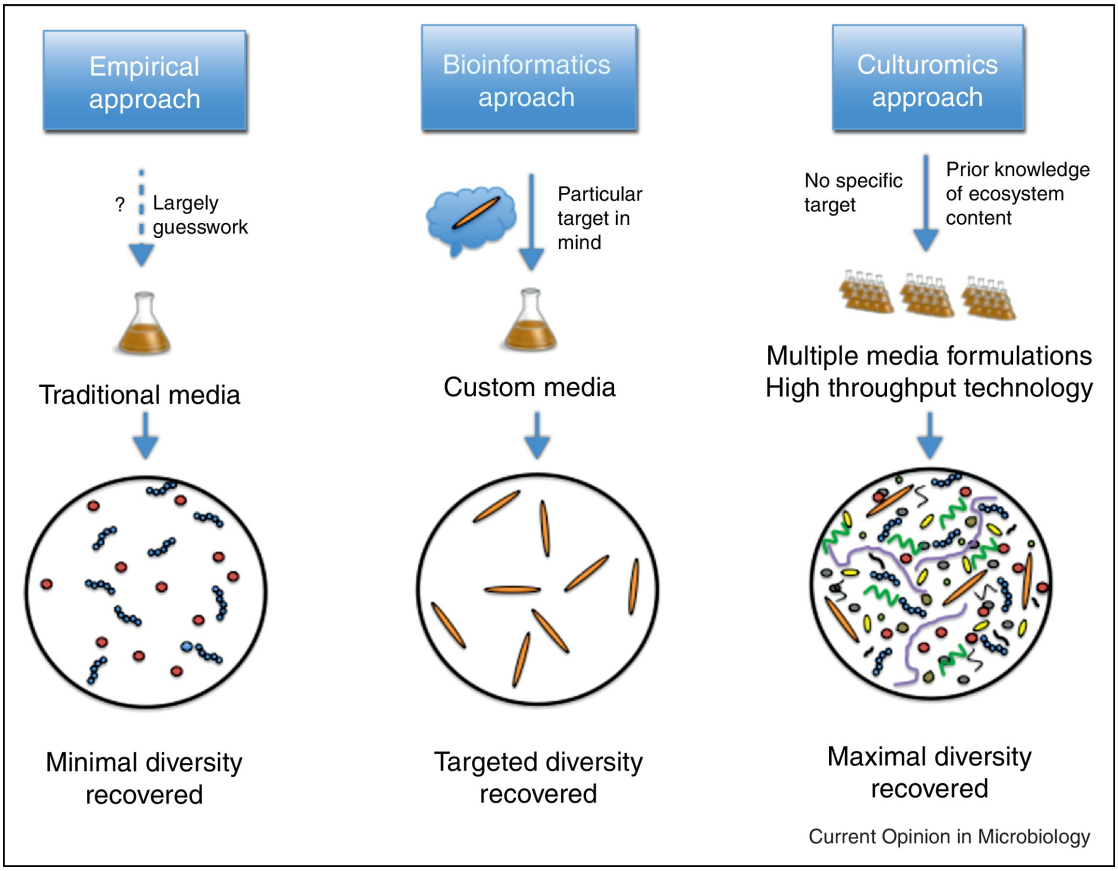


Figure 22: Schematic outlining of the general approaches to culturing microbial diversity

(Allen-Vercoe, 2013)

6.4 Integrative Ecosystems Biology

The present thesis focused on microbial community analyses involving HT sequencing technologies, which were based on DNA and therefore cannot distinguish between metabolically (in)active and dead microorganisms. Important functional information of the metabolically active fraction of a community can be gained from the fields of metatranscriptomics and metaproteomics. These have the inherent advantage to reveal functional activity rather than predicted functional capabilities that are derived from (meta)genomic analyses. Combining these approaches with metagenomics and exhaustive cultivation efforts (Allen-Vercoe, 2013) (Figure 22) could provide new insights into relations between members of a microbial community and will be important for designing interventions targeted at functions of the microbial communities in food rather than specific constituent species. Furthermore, metagenome data can be integrated with (meta)-metabolome data (Turnbaugh & Gordon, 2008). Such integrative ecosystems biology studies [*e.g.*, (Shi *et al.*, 2011; Teeling *et al.*, 2012)] will pose new challenges due to their increasing data complexities, in particular with respect to bioinformatics post-processing. To cope with these huge amounts of data, there will be a growing need for automated pipelines enabling bioinformaticians to perform the analysis and to interpret the outcome.

7

LAB Taxonomy in a HT Sequencing Era

7.1 Polyphasic Taxonomy

Polyphasic taxonomy of LAB comprises the application of comparative 16S rRNA gene sequence analysis, MLSA of less conserved phylogenetic markers, DNA-DNA hybridization, and phenotypic characterization to provide conclusive evidence about the taxonomic status of an organism (as applied in Sections 4.1, 4.2, and 4.3). Furthermore, the determination of the G+C content is part of the standard description of novel taxa, whereas the analysis of the peptidoglycan structure of the cell wall is more specific for LAB taxonomy. Different aspects make this polyphasic taxonomy approach time- and money-consuming, labor-intensive, and not accessible to all laboratories.

Due to the conserved nature of the 16S rRNA gene in LAB species, MLSA of less conserved genes is often required in LAB species descriptions. One aspect that makes MLSA unsuited for HT applications is the lack of universal primers and, as a consequence, novel primer design and PCR optimization are often required to extend the existing MLSA scheme to several LAB genera (as was put into practice for the genus *Carnobacterium* in Section 4.3). In Section 4.1, amplification of the

rpoA and *atpA* genes proved unsuccessful for many strains of the genus *Weissella* using the primers available, and therefore sequencing was restricted to the *pheS* gene. As a consequence, novel primer design and optimization of the PCR will be mandatory to extend the existing MLSA scheme to the genus *Weissella*.

Next to the issues related to sequencing of the 16S rRNA gene and the construction of an MLSA scheme for LAB, DDH experiments also have several inherent characteristics that make it unsuited to apply in HT. These include the requirement of large quantities of high-quality DNA, the time- and money-consuming and labor-intensive nature, the inability to build databases (Gevers *et al.*, 2005), and the lack of intra- and inter-laboratory reproducibility (Stackebrandt & Ebers, 2006; Rosselló-Mora, 2005).

Furthermore, a range of labor-intensive and time-consuming tests are utilized for the determination of phenotypic characteristics of novel taxa. Phenotypic characteristics that are part of the standard description of LAB taxa commonly include a positive Gram reaction, absence of endospores, oxidase and catalase activity (typically absent), glucose fermentation products (typically lactic acid), carbohydrate fermentation patterns, ratio of D- and L-lactic acid production, hydrolysis of aesculin and arginine, reduction of nitrate, gelatine liquefaction, growth at different temperatures, pH range values and NaCl concentrations, and tolerance to oxygen (Mattarelli *et al.*, 2014). Unfortunately, the tests to determine these phenotypic characteristics are time-consuming and labor-intensive and most of these phenotypic characteristics are poorly discriminatory among species and genera and mainly have a descriptive value in LAB taxonomy. Another phenotypic characteristic recommended for novel LAB species descriptions is the peptidoglycan type, which is species-specific for many LAB species. Nevertheless, the diversity in peptidoglycan type is very limited among species, with the Lys-D-Asp type being the predominant type within the genus *Lactobacillus*, *Pediococcus*, and several other LAB genera. The determination of the peptidoglycan type is not suited for HT applications because it is labor-intensive, time- and money-consuming, and not accessible to all laboratories.

As discussed above, the polyphasic taxonomy approach to describe novel LAB taxa is not suited to perform with HT capacity, which is contradictory with the huge amount of strains that are currently awaiting description and formal naming (Vandamme & Peeters, 2014). Furthermore, this approach contrasts with the emerging culturomics approaches, which will generate a huge amount of not yet characterized strains. Therefore, the traditionally used polyphasic taxonomy approach is outdated and should urgently be reconsidered to enable fast description of novel LAB taxa.

7.2 Genomic Taxonomy

With the advent of HT sequencing technologies that are becoming cheaper and the development of many tools that guide bioinformaticians during the assembly and annotation of whole-genome sequences, routine sequencing and downstream analysis come into reach of many research groups. Because whole-genome sequences provide insights into the genetic nature of microbial species, they can be used as a tool for delineating bacterial species and for studying their phylogeny.

LAB genome sequences are continuously accumulating, providing a good opportunity to study the evolutionary history of LAB. For instance, sequences of the 16S rRNA gene and less conserved genes for MLSA can be extracted out of the whole-genome sequence, evading current issues related to primer universality and PCR optimizations (as discussed above). Furthermore, whole-genome sequence-based approaches are superior for studying phylogeny, because of the better resolution of whole-genome sequences for discriminating both distantly and closely related bacteria compared to single or multiple genes (Vandamme & Peeters, 2014). In addition, the G+C content can be calculated from the whole-genome sequence, making the application of HPLC approaches obviate for G+C content analyses.

While the DDH species threshold is being translated into MLSA or whole-genome based thresholds (Gevers *et al.*, 2005; Goris *et al.*, 2007; Richter & Rossello-Mora, 2009; Tindall *et al.*, 2010), the reassessment of bacterial taxonomy and the species definition will require more than a methodological translation of threshold levels (Achtman & Wagner, 2008). In 1987, the ad hoc committee on reconciliation of approaches to bacterial systematics (Wayne *et al.*, 1987) stated that taxonomy should be determined phylogenetically and that the complete genome sequence should therefore be the standard for species delineation. As DDH experiments were originally developed because whole-genome sequences were not available, the current accumulation of whole-genome sequences enables to abandon DDH experiments out of the polyphasic taxonomy approach and to establish taxonomic schemes based on evolutionary information contained in whole-genome sequences (Eisen, 2000; Wolf *et al.*, 2001). Comparing whole-genome sequences have many advantages compared to DDH experiments, such as the ability to build databases and to perform these comparisons in HT.

Furthermore, whole-genome sequences yield information on the metabolic potential of a particular strain (as was applied in Section 5.2), which could make extensive phenotypic tests obviate. Nevertheless, to date, complete genome sequences and improved insights from genome annotations have not yet resulted in reliable predictions of metabolic and chemotaxonomic features, but it is likely that this will occur at some point in the future (Oren & Garrity, 2014). However, it is not known if these properties are actually expressed in the phenotype of the organism as it grows in the laboratory or in its natural environment. Therefore, at the time of writing, analysis of whole-genome sequences cannot be considered a proxy for the phenotypic characterization contained in novel species descriptions.

7.3 Integrating Genomics into LAB Taxonomy

There is a strong consensus that prokaryotic classification should be based on a polyphasic taxonomy approach, integrating data from traditional physiological and chemotaxonomic tests with complementary results from genetic methods (Vandamme *et al.*, 1996; Rossello-Mora & Amann, 2001; Stackebrandt *et al.*, 2002; Schleifer, 2009). Recently, Mattarelli and coworkers (2014) have provided the recommended minimal standards for description of new taxa of the genera *Bifidobacterium*, *Lactobacillus*, and related genera. Explicit in their guidelines is an emphasis that novel LAB taxa should be characterized as comprehensive as possible based on phenotypic, genotypic, and ecological characteristics. Nevertheless, how much characterization is considered enough?

Recently, Vandamme and Peeters (2014) proposed to use a whole-genome sequence and a minimal description of phenotypic characteristics as a basic biological identity card that be considered sufficient, cost-effective, and appropriate for species descriptions, which is a workable option for future large-scale descriptions of novel LAB species. In my opinion, novel LAB species descriptions should include a draft genome sequence accompanied with contextual data (*e.g.*, elaborated description of the sampling site, growth requirements, isolation strategies, date of sampling, ...) and data with taxonomic relevance (*e.g.*, the analysis of carbohydrate fermentation patterns and of glucose fermentation products, the ratio of D- and L-lactic acid production, Gram stain, catalase and oxidase activity). Phenotypic characteristics without or with limited taxonomic relevance, (*e.g.*, cell and colony morphology) should be excluded from LAB species descriptions, preventing taxonomists to perform a plethora of time- and money-consuming physiological tests of questionable

taxonomic relevance (Sutcliffe *et al.*, 2012). Furthermore, the LAB draft genome sequences should be subjected to a range of *in silico* calculations, including ANI, G+C content, phylogenetic trees of the core gene set, the prediction of metabolic pathways of peptidoglycan synthesis, ...

These basic biological identity cards should be stored and maintained in high-quality, readily accessible, iterative, and adaptable taxonomic databases (Sutcliffe *et al.*, 2012). The accumulation of these curated data will enable the determination of thresholds for species delineation and provide valuable resources for different research areas. The re-evaluation of how novel taxa are described will undoubtedly demand online publication formats, coupled to a more effective use of online databases to allow the iterative comparison of prokaryotic characteristics. Finally, these online taxonomic databases should consider how to deal with genome sequences of strains that cannot be cultivated, and hence cannot be deposited into culture collections.

8

Future Perspectives

This Chapter provides perspectives on the integration of HT sequencing technologies in future microbial diversity analyses of food ecosystems and outlines the possible role of genomic taxonomy to improve the quality of microbial diversity analyses. Figure 23 proposes a workflow to characterize the microbial ecosystem contained in a food sample of interest by merging the fields of genomic taxonomy, reference database construction and management, (meta)genomics, (meta)transcriptomics, (meta)-metabolomics, and culturomics, which have the potential to collectively provide a comprehensive, integrative analysis of the food ecosystem of interest.

If a food sample with an unknown microbial ecosystem is provided, the starting point is the extensive collection of contextual metadata (as described in Section 6.2) and the sampling of at least three biological and technical replicates, if possible (Figure 23). These replicates are subsequently pre-discovered using a target enrichment strategy, which includes amplification of the bacterial 16S rRNA gene and the fungal ITS region. Following, these amplicons are subjected to HT sequencing using the platform of interest. This pre-discovery step will provide basic knowledge on the ecosystem of interest, which is crucial for subsequent in-depth analyses.

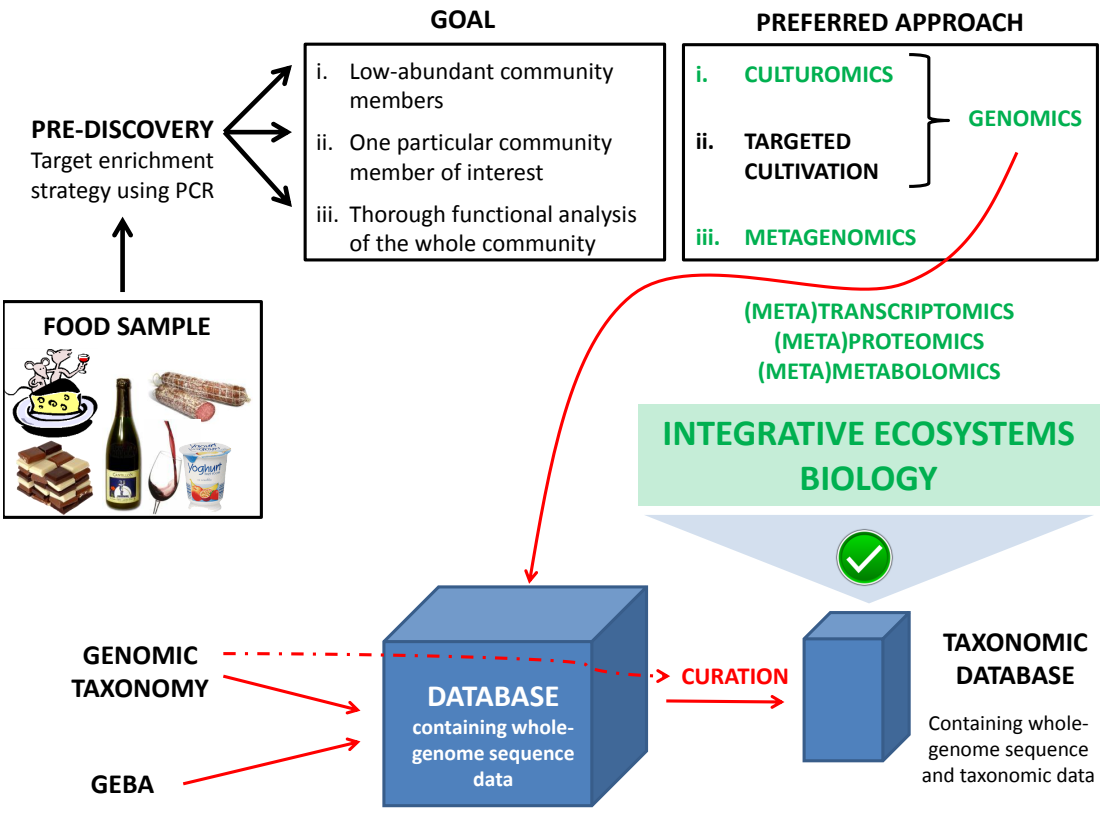


Figure 23: Microbial diversity analyses in a HT sequencing era

This Figure is extensively discussed in Chapter 8

Following, dependent on the outcome of the pre-discovery analysis and the research question, different approaches are preferred. **(i)** If the goal is to obtain an in-depth analysis of the low-abundant community members of the food ecosystem of interest, culturomics is a valuable approach (as discussed in Section 6.3). Knowledge on the identity of these low-abundant community members, which was previously obtained by the pre-discovery approach, is valuable to define a set of growth media and conditions that could promote the growth of these low-abundant community members and could discourage growth of the dominant community members. Robotic plating on an extensive set of growth media and conditions, followed by MALDI-TOF MS dereplication and identification using 16S rRNA gene sequencing, have the potential to significantly increase the throughput of this culturomics approach. **(ii)** If the goal is to target a limited number of community members of interest, the targeted bioinformatics approach is suited to isolate the target strain(s) (as discussed in Section 6.3). Similarly to the culturomics approach, the pre-discovery

step will provide information on the most suited strategy to isolate the target strain(s) of interest. The isolates obtained by the culturomics or the targeted bioinformatics approach [**(i)** & **(ii)**] can subsequently be subjected to whole-genome sequencing. In case of the inability to cultivate the target strain, the whole-genome sequence of that particular strain can be obtained by means of single-cell genomics or metagenome sequencing to infer its necessary growth requirements. These whole-genome sequences add interesting information to reference databases, which can be mined by other researchers. Another value of whole-genome sequencing of bacterial isolates is that functionalities can be assigned to specific microbial lineages, which contrasts with the lack of connectivity (*i.e.*, which species harbors which functionality) of metagenomic datasets (Walker *et al.*, 2014). As such, culture and phylogeny will continue to have crucially important roles in food fermentation microbiota research and will be required for the development of novel starter cultures. **(iii)** If one wants to provide a thorough functional analysis of the whole microbial community, metagenome sequencing is the approach preferred. The previously obtained knowledge on the species richness and composition will assist in determining how deep one needs to sequence to obtain enough coverage.

As (meta)genomics cannot distinguish between metabolically (in)active and dead microorganisms, the fields of (meta)transcriptomics and (meta)proteomics can reveal functional activity rather than predicted functional capabilities, derived from (meta)genomic analyses. Furthermore, combining (meta)genomics, (meta)transcriptomics, (meta)proteomics, and (meta)metabolomics with culturomics could provide new insights into relations between members of a microbial community and will be important for designing interventions targeted at functions of the microbial communities in food rather than specific constituent species. Automated pipelines will enable bioinformaticians to cope with the huge amounts of data and to analyze and interpret the outcome of these integrative ecosystem biology studies.

Although numerous large-scale genome sequencing studies have been performed, they have tended to focus on particular habitats of high interest to the scientific community or on the relatives of specific microorganisms. As a consequence, the currently available whole-genome sequences are limited by their highly biased phylogenetic distribution, creating a major obstacle in progressing towards a genome-based classification of microorganisms. Therefore, the Genomic encyclopaedia of bacteria and archaea (GEBA) project was launched aiming at reducing this bias (Klenk & Goker, 2010). They initiated a phylogenetically driven genome sequencing effort, selecting microorganisms based on their position in a phylogenetic tree of small subunit rRNA gene sequences. Although the GEBA project is already a

first attempt to correct for this bias, taxonomists could contribute by making the availability of a whole-genome sequence a prerequisite for the description of novel bacterial taxa.

Next to reducing the biased phylogenetic distribution of the whole-genome sequences that are currently available, switching to a genomic taxonomy enables the construction, curation, and maintenance of comprehensive taxonomic databases containing valuable contextual and taxonomic data, next to whole-genome sequences. This effort could aid in gene prediction of (meta)genomic datasets, because many gene finders require reference sequences from single species that are subsequently used to build a species-specific gene prediction model (Teeling *et al.*, 2012) (as discussed above). This will also have a major influence on the annotations of (meta)transcriptomic, (meta)proteomic, and (meta)-metabolomic data, which could improve the quality of the integrative ecosystem biology studies, and thereby bring future microbial diversity analyses of food ecosystems to a higher level. Furthermore, these curated taxonomic reference databases can be mined for strains with interesting functionalities, which could be of interest to the food industry.

Part V

Bibliography

Abbott, D. A., Hynes, S. H., & Ingledew, W. M. (2005). Growth rates of *Dekkera/Brettanomyces* yeasts hinder their ability to compete with *Saccharomyces cerevisiae* in batch corn mash fermentations. *Applied Microbiology and Biotechnology*, *66*, 641–647.

Achtman, M., & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, *6*, 431–440.

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *Journal of Bacteriology*, *186*, 2629–2635.

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, *71*, 8966–8969.

Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Richmond, T. A., Middle, C. M., Rodesch, M. J., Packard, C. J., Weinstock, G. M., & Gibbs, R. A. (2007). Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, *4*, 903–905.

Alegria, A., Szczesny, P., Mayo, B., Bardowski, J., & Kowalczyk, M. (2012). Biodiversity in Oscypek, a traditional Polish cheese, determined by culture-dependent and -independent approaches. *Applied and Environmental Microbiology*, *78*, 1890–1898.

Allen-Vercoe, E. (2013). Bringing the gut microbiota into focus through microbial culture: recent progress and future perspective. *Current Opinion in Microbiology*, *16*, 625–9.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*, 3389–3402.

Andorra, I., Monteiro, M., Esteve-Zarzoso, B., Albergaria, H., & Mas, A. (2011). Analysis and direct quantification of *Saccharomyces cerevisiae* and *Hanseniaspora guilliermondii* populations during alcoholic fermentation by fluorescence *in situ* hybridization, flow cytometry and quantitative PCR. *Food Microbiology*, *28*, 1483–1491.

Andres-Barrao, C., Benagli, C., Chappuis, M., Ortega Perez, R., Tonolla, M., & Barja, F. (2013). Rapid identification of acetic acid bacteria using MALDI-TOF mass spectrometry fingerprinting. *Systematic and Applied Microbiology*, *36*, 75–81.

Angiuoli, S. V., Gussman, A., Klimke, W., Cochrane, G., Field, D., Garrity, G., Kodira, C. D., Kyrpides, N., Madupu, R., Markowitz, V., Tatusova, T., Thomson, N., & White, O. (2008). Toward an online repository of standard operating procedures (SOPs) for (meta)genomic annotation. *OMICS*, *12*, 137–141.

Arena, M. P., Romano, A., Capozzi, V., Beneduce, L., Ghariani, M., Grieco, F., Lucas, P., & Spano, G. (2011). Expression of *Lactobacillus brevis* IOEB 9809 tyrosine decarboxylase and agmatine deiminase genes in wine correlates with substrate availability. *Letters in Applied Microbiology*, *53*, 395–402.

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., & Weightman, A. J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, *71*, 7724–7736.

Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., Formsma, K., Gerdes, S., Glass, E., Kubal, M., Meyer, F., Olsen, G., Olson, R., Osterman, A., Overbeek, R., McNeil, L., Paarmann, D., Paczian, T., Parrello, B., Pusch, G., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., & Zagnitko, O. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics*, *9*, 1–15.

Babot, J. D., Hidalgo, M., Arganaraz-Martinez, E., Apella, M. C., & Perez Chaia, A. (2011). Fluorescence *in situ* hybridization for detection of classical propionibacteria with specific 16S rRNA-targeted probes and its application to enumeration in Gruyere cheese. *International Journal of Food Microbiology*, *145*, 221–228.

Begerow, D., Nilsson, H., Unterseher, M., & Maier, W. (2010). Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Microbiology and Biotechnology*, *87*, 99–108.

Behr, J., Israel, L., Gänzle, M. G., & Vogel, R. F. (2007). Proteomic approach for characterization of hop-inducible proteins in *Lactobacillus brevis*. *Applied and Environmental Microbiology*, *73*, 3300–3306.

Behr, J., & Vogel, R. F. (2010). Mechanisms of hop inhibition include the transmembrane redox reaction. *Applied and Environmental Microbiology*, *76*, 142–149.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*, D36–42.

Bergsveinson, J., Pittet, V., & Ziola, B. (2012). RT-qPCR analysis of putative beer-spoilage gene expression during growth of *Lactobacillus brevis* BSO 464 and *Pediococcus claussenii* ATCC BAA-344^{T} in beer. *Applied Microbiology and Biotechnology*, *96*, 461–470.

Bermúdez-Humarán, L. G., Aubry, C., Motta, J.-P., Deraison, C., Steidler, L., Vergnolle, N., Chatel, J.-M., & Langella, P. (2013). Engineering lactococci and lactobacilli for human health. *Current Opinion in Microbiology*, *16*, 278–283.

Bernal, A., Ear, U., & Kyrpides, N. (2001). Genomes onLine database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Research*, *29*, 126–127.

Berry, D., Mahfoudh, K. B., Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, *77*, 7846–7849.

Björkroth, K. J., Schillinger, U., Geisen, R., Weiss, N., Hoste, B., Holzapfel, W. H., Korkeala, H. J., & Vandamme, P. (2002). Taxonomic study of *Weissella confusa* and description of *Weissella cibaria* sp. nov., detected in food and clinical samples. *International Journal of Systematic and Evolutionary Microbiology*, *52*, 141–148.

Blasco, L., Ferrer, S., & Pardo, I. (2003). Development of specific fluorescent oligonucleotide probes for *in situ* identification of wine lactic acid bacteria. *FEMS Microbiology Letters*, *225*, 115–123.

Bokulich, N. A., & Bamforth, C. W. (2013). The microbiology of malting and brewing. *Microbiology and Molecular Biology Reviews*, *77*, 157–172.

Bokulich, N. A., Bamforth, C. W., & Mills, D. A. (2012a). Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale. *PLoS ONE*, *7*, e35507.

Bokulich, N. A., Joseph, C. M., Allen, G., Benson, A. K., & Mills, D. A. (2012b). Next-generation sequencing reveals significant bacterial diversity of botrytized wine. *PLoS ONE*, *7*, e36357.

Bokulich, N. A., & Mills, D. A. (2012c). Differentiation of mixed lactic acid bacteria communities in beverage fermentations using targeted terminal restriction fragment length polymorphism. *Food Microbiology*, *31*, 126–132.

Bokulich, N. A., & Mills, D. A. (2012d). Next-generation approaches to the microbial ecology of food fermentations. *BMB Reports*, *45*, 377–389.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D., & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Research*, *11*, 731–753.

Bottari, B., Santarelli, M., Neviani, E., & Gatti, M. (2010). Natural whey starter for Parmigiano Reggiano: culture-independent approach. *Journal of Applied Microbiology*, *108*, 1676–1684.

Breidt, F., Medina, E., Wafa, D., Pérez-Díaz, I., Franco, W., Huang, H.-Y., Johanningsmeier, S. D., & Kim, J. H. (2013). Characterization of cucumber fermentation spoilage bacteria by enrichment culture and 16S rDNA cloning. *Journal of Food Science*, *78*, M470–M476.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: the artemis comparison tool. *Bioinformatics (Oxford, England)*, *21*, 3422–3423.

Case, R. J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W. F., & Kjelleberg, S. (2007). Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology*, *73*, 278–288.

Chambel, L., Chelo, I. M., Ze-Ze, L., Pedro, L. G., Santos, M. A., & Tenreiro, R. (2006). *Leuconostoc pseudoficulneum* sp. nov., isolated from a ripe fig. *International Journal of Systematic and Evolutionary Microbiology*, *56*, 1375–1381.

Chen, T., Yu, W.-H., Izard, J., Baranova, O. V., Lakshmanan, A., & Dewhirst, F. E. (2010). The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database*, (p. baq013).

Cho, J. C., & Tiedje, J. M. (2001). Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Applied and Environmental Microbiology*, *67*, 3677–3682.

Cho, R. J., Mindrinos, M., Richards, D. R., Sapolsky, R. J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T. L., Stammers, M., Federspiel, N., Theologis, A., Yang, W.-H., Hubbell, E., Au, M., Chung, E. Y., Lashkari, D., Lemieux, B., Dean, C., Lipshutz, R. J., Ausubel, F. M., Davis, R. W., & Oefner, P. J. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics*, *23*, 203–207.

Claesson, M. J., van Sinderen, D., & O'Toole, P. W. (2008). *Lactobacillus* phylogenomics - towards a reclassification of the genus. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 2945–2954.

Clark, C. G., Kruczkiewicz, P., Guan, C., McCorrister, S. J., Chong, P., Wylie, J., van Caeseele, P., Tabor, H. A., Snarr, P., Gilmour, M. W., Taboada, E. N., & Westmacott, G. R. (2013). Evaluation of MALDI-TOF mass spectroscopy methods for determination of *Escherichia coli* pathotypes. *Journal of Microbiological Methods*, *94*, 180–191.

Cleenwerck, I., De Vos, P., & De Vuyst, L. (2010). Phylogeny and differentiation of species of the genus *Gluconacetobacter* and related taxa based on multilocus sequence analyses of housekeeping genes and reclassification of *Acetobacter xylinus* subsp. *sucrofermentans* as *Gluconacetobacter sucrofermentans*. *International Journal of Systematic and Evolutionary Microbiology*, *60*, 2277–2283.

Cocolin, L., Alessandria, V., Dolci, P., Gorra, R., & Rantsiou, K. (2013). Culture independent methods to assess the diversity and dynamics of microbiota during food fermentation. *International Journal of Food Microbiology*, *167*, 29 – 43.

Coenye, T., Falsen, E., Vancanneyt, M., Hoste, B., Govan, J. R. W., Kersters, K., & Vandamme, P. (1999). Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *International Journal of Systematic Bacteriology*, *49*, 405–413.

Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P., & Swings, J. (2005). Towards a prokaryotic genomic taxonomy. *FEMS Microbiology Reviews*, *29*, 147–167.

Coenye, T., & Vandamme, P. (2003). Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case. *Microbiology*, *149*, 3507–3517.

Collins, M. D., Samelis, J., Metaxopoulos, J., & Wallbanks, S. (1993). Taxonomic studies on some *Leuconostoc*-like organisms from fermented sausages: description of a new genus *Weissella* for the *Leuconostoc paramesenteroides* group of species. *Journal of Applied Microbiology*, *75*, 595–603.

Coorevits, A., De Jonghe, V., Vandroemme, J., Reekmans, R., Heyrman, J., Messens, W., De Vos, P., & Heyndrickx, M. (2008). Comparative analysis of the diversity of aerobic spore-forming bacteria in raw milk from organic and conventional dairy farms. *Systematic and Applied Microbiology*, *31*, 126–140.

de Crecy-Lagard, V., El Yacoubi, B., de la Garza, R. D., Noiriel, A., & Hanson, A. D. (2007). Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genomics*, *8*, 245.

Crous, P. W., Gams, W., Stalpers, J. A., Robert, V., & Stegehuis, G. (2004). MycoBank: an online initiative to launch mycology into the 21st century. *Studies in Mycology*, *50*, 19–22.

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). MAUVE: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, *14*, 1394–1403.

Darmon, E., & Leach, D. R. F. (2014). Bacterial genome instability. *Microbiology and Molecular Biology Reviews*, *78*, 1–39.

De Bruyne, K., Camu, N., De Vuyst, L., & Vandamme, P. (2010). *Weissella fabaria* sp. nov., from a Ghanaian cocoa fermentation. *International Journal of Systematic and Evolutionary Microbiology*, *60*, 1999–2005.

De Bruyne, K., Camu, N., Lefebvre, K., De Vuyst, L., & Vandamme, P. (2008a). *Weissella ghanensis* sp. nov., isolated from a Ghanaian cocoa fermentation. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 2721–2725.

De Bruyne, K., Franz, C. M., Vancanneyt, M., Schillinger, U., Mozzi, F., de Valdez, G. F., De Vuyst, L., & Vandamme, P. (2008b). *Pediococcus argentinicus* sp. nov. from Argentinean fermented wheat flour and identification of *Pediococcus* species by *pheS*, *rpoA* and *atpA* sequence analysis. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 2909–2916.

De Bruyne, K., Schillinger, U., Caroline, L., Boehringer, B., Cleenwerck, I., Vancanneyt, M., De Vuyst, L., Franz, C. M. A. P., & Vandamme, P. (2007). *Leuconostoc holzapfelii* sp. nov., isolated from Ethiopian coffee fermentation and assessment of sequence analysis of housekeeping genes for delineation of *Leuconostoc* species. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 2952–2959.

De Bruyne, K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B., & Vandamme, P. (2011). Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning. *Systematic and Applied Microbiology*, *34*, 20–29.

De Clerck, E., & De Vos, P. (2002). Study of the bacterial load in a gelatine production process focussed on *Bacillus* and related endosporeforming genera. *Systematic and Applied Microbiology*, *25*, 611–617.

De Vuyst, L., Lefeber, T., Papalexandratou, Z., & Camu, N. (2010). The functional role of lactic acid bacteria in cocoa bean fermentation. In *Biotechnology of lactic acid bacteria* (pp. 301–325). Wiley-Blackwell.

Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, *23*, 673–679.

Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research*, *27*, 4636–4641.

Delmont, T. O., Malandain, C., Prestat, E., Larose, C., Monier, J.-M., Simonet, P., & Vogel, T. M. (2011). Metagenomic mining for microbiologists. *ISME Journal*, *5*, 1837–1843.

Deloger, M., El Karoui, M., & Petit, M. A. (2009). A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *Journal of Bacteriology*, *191*, 91–99.

DeLong, E. F., & Karl, D. M. (2005). Genomic perspectives in microbial oceanography. *Nature*, *437*, 336–342.

Derrien, M., Vaughan, E. E., Plugge, C. M., & de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *54*, 1469–1476.

Di Cagno, R., Coda, R., De Angelis, M., & Gobbetti, M. (2013). Exploitation of vegetables and fruits through lactic acid fermentation. *Food Microbiology*, *33*, 1–10.

Di Mattia, E., Grego, S., & Cacciari, I. (2002). Eco-physiological characterization of soil bacterial populations in different states of growth. *Microbial Ecology*, *43*, 34–43.

Dickie, I. A. (2010). Insidious effects of sequencing errors on perceived diversity in molecular surveys. *New Phytologist*, *188*, 916–918.

Dieckmann, R., Graeber, I., Kaesler, I., Szewzyk, U., & von Dohren, H. (2005). Rapid screening and dereplication of bacterial isolates from marine sponges of the sula ridge by intact-cell-MALDI-TOF mass spectrometry (ICM-MS). *Applied Microbiology and Biotechnology*, *67*, 539–548.

Dijkshoorn, L., Ursing, B. M., & Ursing, J. B. (2000). Strain, clone and species: comments on three basic concepts of bacteriology. *Journal of Medical Microbiology*, *49*, 397–401.

Doolittle, W. F., & Papke, R. T. (2006). Genomics and the bacterial species problem. *Genome Biology*, *7*, 116.

Duskova, M., Sedo, O., Ksicova, K., Zdrahal, Z., & Karpiskova, R. (2012). Identification of lactobacilli isolated from food by genotypic methods and MALDI-TOF MS. *International Journal of Food Microbiology*, *159*, 107–114.

Dykhuizen, D. (2005). Species numbers in bacteria. *Proceedings of the California Academy of Sciences*, *56*, 62–71.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, *14*, 755–763.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*, 2194–2200.

Ehrmann, M. A., Freiding, S., & Vogel, R. F. (2009). *Leuconostoc palmae* sp. nov., a novel lactic acid bacterium isolated from palm wine. *International Journal of Systematic and Evolutionary Microbiology*, *59*, 943–947.

Eisen, J. A. (2000). Assessing evolutionary relationships among microbes from whole-genome analysis. *Current Opinion in Microbiology*, *3*, 475–480.

Endo, A., Irisawa, T., Futagawa-Endo, Y., Sonomoto, K., Itoh, K., Takano, K., Okada, S., & Dicks, L. M. T. (2010). *Fructobacillus tropaeoli* sp. nov., a novel fructophilic lactic acid bacterium isolated from a flower. *International Journal of Systematic and Evolutionary Microbiology*, (pp. ijs.0.023838–0).

Endo, A., & Okada, S. (2008). Reclassification of the genus *Leuconostoc* and proposals of *Fructobacillus fructosus* gen. nov., comb. nov., Fructobacillus durionis comb. nov., *Fructobacillus ficulneus* comb. nov. and *Fructobacillus pseudoficulneus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *58*, 2195–2205.

Engelbrektson, A., Kunin, V., Wrighton, K. C., Zvenigorodsky, N., Chen, F., Ochman, H., & Hugenholtz, P. (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME Journal*, *4*, 642–647.

Ercolini, D. (2013). High-throughput sequencing and metagenomics: Moving forward in the culture-independent analysis of food microbial ecology. *Applied and Environmental Microbiology*, *79*, 3148–3155.

Ezaki, T., Hashimoto, Y., & Yabuuchi, E. (1989). Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. *International Journal of Systematic Bacteriology*, *39*, 224–229.

Fernandez, M., del Rio, B., Linares, D. M., Martin, M. C., & Alvarez, M. A. (2006). Real-time polymerase chain reaction for quantitative detection of histamine-producing bacteria: use in cheese production. *Journal of Dairy Science*, *89*, 3763–3769.

Field, D., Amaral-Zettler, L., Cochrane, G., Cole, J. R., Dawyndt, P., Garrity, G. M., Gilbert, J., Glockner, F. O., Hirschman, L., Karsch-Mizrachi, I., Klenk, H. P., Knight, R., Kottmann, R., Kyrpides, N., Meyer, F., San Gil, I., Sansone, S. A., Schriml, L. M., Sterk, P., Tatusova, T., Ussery, D. W., White, O., & Wooley, J. (2011). The genomic standards consortium. *PLoS Biology*, *9*, e1001088.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., Allen, M. J., Angiuoli, S. V., Ashburner, M., Axelrod, N., Baldauf, S., Ballard, S., Boore, J., Cochrane, G., Cole, J., Dawyndt, P., De Vos, P., DePamphilis, C., Edwards, R., Faruque, N., Feldman, R., Gilbert, J., Gilna, P., Glockner, F. O., Goldstein, P., Guralnick, R., Haft, D., Hancock, D., Hermjakob, H., Hertz-Fowler, C., Hugenholtz, P., Joint, I., Kagan, L., Kane, M., Kennedy, J., Kowalchuk, G., Kottmann, R., Kolker, E., Kravitz, S., Kyrpides, N., Leebens-Mack, J., Lewis, S. E., Li, K., Lister, A. L., Lord, P., Maltsev, N., Markowitz, V., Martiny, J., Methe, B., Mizrachi, I., Moxon, R., Nelson, K., Parkhill, J., Proctor, L., White, O., Sansone, S. A., Spiers, A., Stevens, R., Swift, P., Taylor, C., Tateno, Y., Tett, A., Turner, S., Ussery, D., Vaughan, B., Ward, N., Whetzel, T., San Gil, I., Wilson, G., & Wipat, A. (2008). The minimum information about a genome sequence (MIGS) specification. *Nature Biotechnology*, *26*, 541–547.

Fleet, G. H. (1999). Microorganisms in food ecosystems. *International Journal of Food Microbiology*, *50*, 101–117.

Fleet, G. H. (2007). Yeasts in foods and beverages: impact on product quality and safety. *Current Opinion in Biotechnology*, *18*, 170–175.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., & al., E. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, *269*, 496–512.

Franzmann, P. D., Höpfl, P., Weiss, N., & Tindall, B. J. (1991). Psychrotrophic, lactic acid-producing bacteria from anoxic waters in Ace Lake, Antarctica; *Carnobacterium funditum* sp. nov. and *Carnobacterium alterfunditum* sp. nov. *Archives of Microbiology*, *156*, 255–262.

Fraser, C., Hanage, W. P., & Spratt, B. G. (2007). Recombination and the nature of bacterial speciation. *Science*, *315*, 476–480.

Freiwald, A., & Sauer, S. (2009). Phylogenetic classification and identification of bacteria by mass spectrometry. *Nature Protocols*, *4*, 732–742.

Garcia-Armisen, T., Papalexandratou, Z., Hendryckx, H., Camu, N., Vrancken, G., Vuyst, L., & Cornelis, P. (2010). Diversity of the total bacterial community associated with Ghanaian and Brazilian cocoa bean fermentation samples as revealed by a 16 S rRNA gene clone library. *Applied Microbiology and Biotechnology*, *87*, 2281–2292.

García-Gimeno, R. M., & Zurera-Cosano, G. (1997). Determination of ready-to-eat vegetable salad shelf-life. *International Journal of Food Microbiology*, *36*, 31–38.

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for *basidiomycetes* - application to the identification of mycorrhizae and rusts. *Molecular Ecology*, *2*, 113–118.

Garrity, G. M., Bell, J. A., & Lilburn, T. G. (2004). *Taxonomic outline of the prokaryotes. Bergey's manual of systematic bacteriology.*. New York: Springer.

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F. L., & Swings, J. (2005). Opinion: Re-evaluating prokaryotic species. *Nature Reviews Microbiology*, *3*, 733–739.

Gevers, D., Huys, G., & Swings, J. (2001). Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. *FEMS Microbiology Letters*, *205*, 31–36.

Ghyselinck, J., Van Hoorde, K., Hoste, B., Heylen, K., & De Vos, P. (2011). Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. *Journal of Microbiological Methods*, *86*, 327–336.

Gkatzionis, K., Yunita, D., Linforth, R. S. T., Dickinson, M., & Dodd, C. E. R. (2014). Diversity and activities of yeasts from different parts of a Stilton cheese. *International Journal of Food Microbiology*, *177*, 109–116.

Goll, J., Rusch, D. B., Tanenbaum, D. M., Thiagarajan, M., Li, K., Methé, B. A., & Yooseph, S. (2010). METAREP: JCVI metagenomics reports-an open source tool for high-performance comparative metagenomics. *Bioinformatics*, *26*, 2631–2632.

Goris, J., Konstantinidis, K., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 81–91.

Goris, J., Suzuki, K.-i., Vos, P. D., Nakase, T., & Kersters, K. (1998). Evaluation of a microplate DNA - DNA hybridization method compared with the initial renaturation method. *Canadian Journal of Microbiology*, *44*, 1148–1153.

Gouy, M., Guindon, S. A., & Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, *27*, 221–224.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., The Human Microbiome, C., Petrosino, J. F., Knight, R., & Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, *21*, 494–504.

Hall, B. (2011). *Phylogenetic trees made easy: A how-to manual*. (4th ed.). Sinauer Associates, Inc. Publishers.

Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E. A., Karlin-Neumann, G. A., Fakhrai-Rad, H., Ronaghi, M., Willis, T. D., Landegren, U., & Davis, R. W. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnology*, *21*, 673–678.

Hartmann, M., Howes, C. G., Abarenkov, K., Mohn, W. W., & Nilsson, R. H. (2010). V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. *Journal of Microbiological Methods*, *83*, 250–253.

Hayashi, N., Ito, M., Horiike, S., & Taguchi, H. (2001). Molecular cloning of a putative divalent-cation transporter gene as a new genetic marker for the identification of *Lactobacillus brevis* strains capable of growing in beer. *Applied Microbiology and Biotechnology*, *55*, 596–603.

van Hijum, S. A., Vaughan, E. E., & Vogel, R. F. (2013). Application of state-of-art sequencing technologies to indigenous food fermentations. *Current Opinion in Biotechnology*, *24*, 178 – 186.

Holdeman, L. V., & Moore, W. E. C. (1977). *Anaerobic laboratory manual*. (4th ed.). Blacksburg, Virginia, USA: Virginia Polytechnic Institute and State University, Anaerobe Laboratory.

Holzapfel, W., Franz, C. A. P., Ludwig, W., Back, W., & Dicks, L. T. (2006). The genera *Pediococcus* and *Tetragenococcus*. In M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, & E. Stackebrandt (Eds.), *The Prokaryotes* chapter 8. (pp. 229–266). Springer US.

Huber, T., Faulkner, G., & Hugenholtz, P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, *20*, 2317–2319.

Humblot, C., & Guyot, J. P. (2009). Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as pearl millet slurries. *Applied and Environmental Microbiology*, *75*, 4354–4361.

Huse, S., Huber, J., Morrison, H., Sogin, M., & Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, *8*, R143.

Huse, S., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, *12*, 1889–1898.

Huson, D., & Mitra, S. (2012). Introduction to the analysis of environmental sequences: metagenomics with MEGAN. In M. Anisimova (Ed.), *Evolutionary genomics* chapter 17. (pp. 415–429). Humana Press volume 856.

Ibarburu, I., Aznar, R., Elizaquivel, P., Garcia-Quintans, N., Lopez, P., Munduate, A., Irastorza, A., & Duenas, M. T. (2010). A real-time PCR assay for detection and quantification of 2-branched (1,3)-beta-D-glucan producing lactic acid bacteria in cider. *International Journal of Food Microbiology*, *143*, 26–31.

Iijima, K., Suzuki, K., Asano, S., Kuriyama, H., & Kitagawa, Y. (2007). Isolation and identification of potential beer-spoilage *Pediococcus inopinatus* and beer-spoilage *Lactobacillus backi* strains carrying the *horA* and *horC* gene clusters. *Journal of the Institute of Brewing*, *113*, 96–101.

Iijima, K., Suzuki, K., Ozaki, K., & Yamashita, H. (2006). HorC confers beer-spoilage ability on hop-sensitive *Lactobacillus brevis* ABBC45cc. *Journal of Applied Microbiology*, *100*, 1282–1288.

Illeghems, K., De Vuyst, L., Papalexandratou, Z., & Weckx, S. (2012). Phylogenetic analysis of a spontaneous cocoa bean fermentation metagenome reveals new insights into its bacterial and fungal community diversity. *PLoS ONE*, *7*, e38040.

Illeghems, K., De Vuyst, L., & Weckx, S. (2013). Complete genome sequence and comparative analysis of *Acetobacter pasteurianus 386B*, a strain well-adapted to the cocoa bean fermentation ecosystem. *BMC Genomics*, *14*, 526.

Irlinger, F. (2008). Safety assessment of dairy microorganisms: coagulase-negative staphylococci. *International Journal of Food Microbiology*, *126*, 302–310.

Janssens, M., Myter, N., De Vuyst, L., & Leroy, F. (2012). Species diversity and metabolic impact of the microbiota are low in spontaneously acidified Belgian sausages with an added starter culture of *Staphylococcus carnosus*. *Food Microbiology*, *29*, 167–177.

Jung, J. Y., Lee, S. H., Kim, J. M., Park, M. S., Bae, J.-W., Hahn, Y., Madsen, E. L., & Jeon, C. O. (2011). Metagenomic analysis of kimchi, a traditional Korean fermented food. *Applied and Environmental Microbiology*, *77*, 2264–2274.

Jung, M. J., Nam, Y. D., Roh, S. W., & Bae, J. W. (2012). Unexpected convergence of fungal and bacterial communities during fermentation of traditional Korean alcoholic beverages inoculated with various natural starters. *Food Microbiology*, *30*, 112–123.

Kalisky, T., & Quake, S. R. (2011). Single-cell genomics. *Nature Methods*, *8*, 311–314.

Kim, J., Chun, J., & Han, H. U. (2000). *Leuconostoc kimchii* sp. nov., a new species from kimchi. *International Journal of Systematic and Evolutionary Microbiology*, *50*, 1915–1919.

Kim, M., & Chun, J. (2005). Bacterial community structure in kimchi, a Korean fermented vegetable food, as revealed by 16S rRNA gene analysis. *International Journal of Food Microbiology*, *103*, 91–96.

Kim, M.-S., Roh, S. W., Nam, Y.-D., Yoon, J.-H., & Bae, J.-W. (2009). *Carnobacterium jeotgali* sp. nov., isolated from a Korean traditional fermented food. *International Journal of Systematic and Evolutionary Microbiology*, *59*, 3168–3171.

Kim, Y. S., Kim, M. C., Kwon, S. W., Kim, S. J., Park, I. C., Ka, J. O., & Weon, H. Y. (2011). Analyses of bacterial communities in meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. *The Journal of Microbiology*, *49*, 340–348.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 111–120.

Kiyohara, M., Koyanagi, T., Matsui, H., Yamamoto, K., Take, H., Katsuyama, Y., Tsuji, A., Miyamae, H., Kondo, T., Nakamura, S., Katayama, T., & Kumagai, H. (2012). Changes in microbiota population during fermentation of narezushi as revealed by pyrosequencing analysis. *Bioscience, Biotechnology, and Biochemistry*, *76*, 48–52.

Klappenbach, J. A., Dunbar, J. M., & Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, *66*, 1328–1333.

Klenk, H. P., & Goker, M. (2010). En route to a genome-based classification of Archaea and Bacteria? *Systematic and Applied Microbiology*, *33*, 175–182.

Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J., Hugenholtz, P., van der Lelie, D., Meyer, F., Stevens, R., Bailey, M., Gordon, J., Kowalchuk, G., & Gilbert, J. (2012). Unlocking the potential of metagenomics through replicated experimental design. *Nature biotechnology*, *30*, 513âĂŤ520. doi:10.1038/nbt.2235.

Kodama, Y., Shumway, M., Leinonen, R., & International Nucleotide Sequence Database, C. (2012). The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Research*, *40*, D54–6.

Konstantinidis, K., Ramette, A., & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 1929–1940.

Konstantinidis, K., & Stackebrandt, E. (2013). Defining taxonomic ranks. In E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The prokaryotes* chapter 4. (pp. 229–254). Springer Berlin Heidelberg.

Konstantinidis, K., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 2567–2572.

Kottmann, R., Gray, T., Murphy, S., Kagan, L., Kravitz, S., Lombardot, T., Field, D., Glockner, F. O., & Genomic Standards, C. (2008). A standard MIGS/MIMS compliant XML schema: toward the development of the genomic contextual data markup language (GCDML). *OMICS*, *12*, 115–121.

Koyanagi, T., Kiyohara, M., Matsui, H., Yamamoto, K., Kondo, T., Katayama, T., & Kumagai, H. (2011). Pyrosequencing survey of the microbial diversity of 'narezushi', an archetype of modern Japanese sushi. *Letters in Applied Microbiology*, *53*, 635–640.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, .

Kumara, H. M. C. S., De Cort, S., & Verachtert, H. (1993). Localization and characterization of α-glucosidase activity in *Brettanomyces lambicus*. *Applied and Environmental Microbiology*, *59*, 2352–2358.

Kumara, H. M. C. S., & Verachtert, H. (1991). Identification of lambic superattenuating micro-organisms by the use of selective antibiotics. *Journal of the Institute of Brewing*, *97*, 181–185.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews*, *72*, 557–578.

Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, *12*, 118–123.

Kurtzman, C. P., & Robnett, C. J. (1998). Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek*, *73*, 331–371.

Lacerda, I. C. A., Gomes, F. C. O., Borelli, B. M., Faria Jr, C. L. L., Franco, G. R., Mourão, M. M., Morais, P. B., & Rosa, C. A. (2011). Identification of the bacterial community responsible for traditional fermentation during sour cassava starch, cachaça and minas cheese production using culture-independent 16s rRNA gene sequence analysis. *Brazilian Journal of Microbiology*, *42*, 650–657.

Ladero, V., Coton, M., Fernandez, M., Buron, N., Martin, M. C., Guichard, H., Coton, E., & Alvarez, M. A. (2011). Biogenic amines content in Spanish and French natural ciders: application of qPCR for quantitative detection of biogenic amine-producers. *Food Microbiology*, *28*, 554–561.

Lagier, J. C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J. F., Koonin, E. V., La Scola, B., & Raoult, D. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clinical Microbiology and Infection*, *18*, 1185–1193.

Lapage, S., Sneath, P., Lessel, E., Skerman, V., Seeliger, H., & Clark, W. (1992). *International code of nomenclature of bacteria: Bacteriological code (1990 revision). Bacterial Code.*. Washington DC: American Society for Microbiology.

Laureys, D., & De Vuyst, L. (2014). Microbial species diversity, community dynamics, and metabolite kinetics of water kefir fermentation. *Applied and Environmental Microbiology*, *80*, 2564–2572.

Lee, J. S., Lee, K. C., Ahn, J.-S., Mheen, T.-I., Pyun, Y.-R., & Park, Y.-H. (2002). *Weissella koreensis* sp. nov., isolated from kimchi. *International Journal of Systematic and Evolutionary Microbiology*, *52*, 1257–1261.

Lee, S. H., Park, M. S., Jung, J. Y., & Jeon, C. O. (2012). *Leuconostoc miyukkimchii* sp. nov., isolated from brown algae (*Undaria pinnatifida*) kimchi. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 1098–1103.

Lefébure, T., & Stanhope, M. (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology*, *8*, 1–17.

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tarraga, A., Cheng, Y., Cleland, I., Faruque, N., Goodgame, N., Gibson, R., Hoad, G., Jang, M., Pakseresht, N., Plaister, S., Radhakrishnan, R., Reddy, K., Sobhany, S., Ten Hoopen, P., Vaughan, R., Zalunin, V., & Cochrane, G. (2011). The European nucleotide archive. *Nucleic Acids Research*, *39*, D28–31.

Leroy, F., & De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science and Technology*, *15*, 67–78.

Leroy, F., Verluyten, J., & De Vuyst, L. (2006). Functional meat starter cultures for improved sausage fermentation. *International Journal of Food Microbiology*, *106*, 270–285.

Li, X. R., Ma, E. B., Yan, L. Z., Meng, H., Du, X. W., Zhang, S. W., & Quan, Z. X. (2011). Bacterial and fungal diversity in the traditional Chinese liquor fermentation process. *International Journal of Food Microbiology*, *146*, 31–37.

Livermore, J. A., Emrich, S. J., Tan, J., & Jones, S. E. (2014). Freshwater bacterial lifestyles inferred from comparative genomics. *Environmental Microbiology*, *16*, 746–758.

Lodolo, E. J., Kock, J. L. F., Axcell, B. C., & Brooks, M. (2008). The yeast *Saccharomyces cerevisiae*-the main character in beer brewing. *FEMS Yeast Research*, *8*, 1018–1036.

Loman, N. J., Constantinidou, C., Chan, J. Z. M., Halachev, M., Sergeant, M., Penn, C. W., Robinson, E. R., & Pallen, M. J. (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*, *10*, 599–606.

Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, *73*, 1576–1585.

Lucas, P. M., Claisse, O., & Lonvaud-Funel, A. (2008). High frequency of histamine-producing bacteria in the enological environment and instability of the histidine decarboxylase production phenotype. *Applied and Environmental Microbiology*, *74*, 811–817.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., & Others, . (2004). ARB: a software environment for sequence data. *Nucleic Acids Research*, *32*, 1363–1371.

Lukashin, A. V., & Borodovsky, M. (1998). GeneMark.HMM: new solutions for gene finding. *Nucleic Acids Research*, *26*, 1107–1115.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., & Konstantinidis, K. T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE*, *7*, e30087.

Lyhs, U., Koort, J. M. K., Lundström, H. S., & Björkroth, K. J. (2004). *Leuconostoc gelidum* and *Leuconostoc gasicomitatum* strains dominated the lactic acid bacterium population associated with strong slime formation in an acetic-acid herring preserve. *International Journal of Food Microbiology*, *90*, 207–218.

Magnusson, J., Jonsson, H., Schnürer, J., & Roos, S. (2002). *Weissella soli* sp. nov., a lactic acid bacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, *52*, 831–834.

Mahenthiralingam, E., Campbell, M. E., Henry, D. A., & Speert, D. P. (1996). Epidemiology of *Burkholderia cepacia* infection in patients with cystic fibrosis: analysis by randomly amplified polymorphic DNA fingerprinting. *Journal of Clinical Microbiology*, *34*, 2914–2920.

Maidak, B. L., Olsen, G. J., Larsen, N., Overbeek, R., McCaughey, M. J., & Woese, C. R. (1997). The RDP (ribosomal database project). *Nucleic Acids Research*, *25*, 109–111.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., Shakhova, V., Grigoriev, I., Lou, Y., Rohksar, D., Lucas, S., Huang, K., Goodstein, D. M., Hawkins, T., Plengvidhya, V., Welker, D., Hughes, J., Goh, Y., Benson, A., Baldwin, K., Lee, J. H., Diaz-Muniz, I., Dosti, B., Smeianov, V., Wechter, W., Barabote, R., Lorca, G., Altermann, E., Barrangou, R., Ganesan, B., Xie, Y., Rawsthorne, H., Tamir, D., Parker, C., Breidt, F., Broadbent, J., Hutkins, R., O'Sullivan, D., Steele, J., Unlu, G., Saier, M., Klaenhammer, T., Richardson, P., Kozyavkin, S., Weimer, B., & Mills, D. (2006). Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 15611–15616.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*, 111–118.

Marcobal, A., Underwood, M. A., & Mills, D. A. (2008). Rapid determination of the bacterial composition of commercial probiotic products by terminal restriction fragment length polymorphism analysis. *Journal of Pediatric Gastroenterology and Nutrition*, *46*, 608–611.

Markowitz, V. M., Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., Pagani, I., Tringe, S., Huntemann, M., Billis, K., Varghese, N., Tennessen, K., Mavromatis, K., Pati, A., Ivanova, N. N., & Kyrpides, N. C. (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Research*, *42*, D568–D573.

Martens, H., Iserentant, D., & Verachtert, H. (1997). Microbiological aspects of a mixed yeast-bacterial fermentation in the production of a special Belgian acidic ale. *Journal of the Institute of Brewing*, *103*, 85–91.

Martín-Platero, A. M., Maqueda, M., Valdivia, E., Purswani, J., & Martínez-Bueno, M. (2009). Polyphasic study of microbial communities of two Spanish farmhouse goats' milk cheeses from Sierra de Aracena. *Food Microbiology*, *26*, 294–304.

Mattarelli, P., Holzapfel, W., Franz, C. M. A. P., Endo, A., Felis, G. E., Hammes, W., Pot, B., Dicks, L., & Dellaglio, F. (2014). Recommended minimal standards for description of new taxa of the genera *Bifidobacterium*, *Lactobacillus* and related genera. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 1434–1451.

Mavromatis, K., Chu, K., Ivanova, N., Hooper, S. D., Markowitz, V. M., & Kyrpides, N. C. (2009). Gene context analysis in the integrated microbial genomes (IMG) data management system. *PLoS ONE*, *4*, e7979.

Mendes-Ferreira, A., Barbosa, C., Jimenez-Marti, E., Del Olmo, M. L., & Mendes-Faia, A. (2010). The wine yeast strain-dependent expression of genes implicated in sulfide production in response to nitrogen availability. *Journal of Microbiology and Biotechnology*, *20*, 1314–1321.

Mesbah, M., & Whitman, W. B. (1989). Measurement of deoxyguanosine/thymidine ratios in complex mixtures by high-performance liquid chromatography for determination of the mole percentage guanine + cytosine of DNA. *Journal of Chromatography*, *479*, 297–306.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research*, *15*, 1767–1776.

Meuzelaar, L. S., Lancaster, O., Pasche, J. P., Kopal, G., & Brookes, A. J. (2007). MegaPlex PCR: a strategy for multiplex amplification. *Nature Methods*, *4*, 835–837.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., & Edwards, R. A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*, 386.

Midha, S., Ranjan, M., Sharma, V., Kumari, A., Singh, P. K., Korpole, S., & Patil, P. B. (2012). Genome sequence of *Pediococcus pentosaceus* strain IE-3. *Journal of Bacteriology*, *194*, 4468.

Moreno Switt, A. I., Andrus, A. D., Ranieri, M. L., Orsi, R. H., Ivy, R., den Bakker, H. C., Martin, N. H., Wiedmann, M., & Boor, K. J. (2014). Genomic comparison of sporeforming bacilli isolated from milk. *BMC Genomics*, *15*, 26.

Mounier, J., Monnet, C., Jacques, N., Antoinette, A., & Irlinger, F. (2009). Assessment of the microbial diversity at the surface of Livarot cheese using culture-dependent and independent approaches. *International Journal of Food Microbiology*, *133*, 31–37.

Muyzer, G., de Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, *59*, 695–700.

Nakamura, Y., Cochrane, G., Karsch-Mizrachi, I., & International Nucleotide Sequence Database, C. (2013). The international nucleotide sequence database collaboration. *Nucleic Acids Research*, *41*, D21–4.

Nam, Y. D., Lee, S. Y., & Lim, S. I. (2012a). Microbial community analysis of Korean soybean pastes by next-generation sequencing. *International Journal of Food Microbiology*, *155*, 36–42.

Nam, Y. D., Park, S. L., & Lim, S. I. (2012b). Microbial composition of the Korean traditional food "kochujang" analyzed by a massive sequencing technique. *Journal of Food Science*, *77*, M250–6.

Naser, S., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M., & Swings, J. (2007). Identification of lactobacilli by *pheS* and *rpoA* gene sequence analyses. *International Journal of Systematic and Evolutionary Microbiology*, *57*, 2777–2789.

Naser, S., Thompson, F. L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M., & Swings, J. (2005a). Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology*, *151*, 2141–2150.

Naser, S., Thompson, F. L., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Thompson, C. C., Vancanneyt, M., & Swings, J. (2005b). Phylogeny and identification of enterococci by *atpA* gene sequence analysis. *Journal of Clinical Microbiology*, *43*, 2224–2230.

Nesbo, C. L., Dlutek, M., & Doolittle, W. F. (2006). Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics*, *172*, 759–769.

Nguyen, D. T. L., Hoorde, K. V., Cnockaert, M., Brandt, E. D., Bruyne, K. D., Le, B. T., & Vandamme, P. (2013). A culture-dependent and -independent approach for the identification of lactic acid bacteria associated with the production of nem chua, a vietnamese fermented meat product. *Food Research International*, *50*, 232 – 240.

Nilsson, R. H., Veldre, V., Hartmann, M., Unterseher, M., Amend, A., Bergsten, J., Kristiansson, E., Ryberg, M., Jumpponen, A., & Abarenkov, K. (2010). An open source software package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-throughput community assays and molecular ecology. *Fungal Ecology*, *3*, 284–287.

Oren, A., & Garrity, G. (2014). Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie van Leeuwenhoek*, *106*, 43–56.

Padonou, S. W., Schillinger, U., Nielsen, D. S., Franz, C. M. A. P., Hansen, M., Hounhouigan, J. D., Nago, M. C., & Jakobsen, M. (2010). *Weissella beninensis* sp. nov., a motile lactic acid bacterium from submerged cassava fermentations, and emended description of the genus *Weissella*. *International Journal of Food Microbiology*, *60*, 2193–2198.

Panelli, S., Brambati, E., Bonacina, C., & Feligini, M. (2013). Detection of *Clostridium tyrobutyricum* in milk to prevent late blowing in cheese by automated ribosomal intergenic spacer analysis. *Journal of Food Science*, *78*, M1569–M1574.

Pangallo, D., Šaková, N., KoreÅĹová, J., Puškárová, A., Kraková, L., Valík, L., & Kuchta, T. (2014). Microbial diversity and dynamics during the production of May bryndza cheese. *International Journal of Food Microbiology*, *170*, 38–43.

Papalexandratou, Z., Camu, N., Falony, G., & De Vuyst, L. (2011a). Comparison of the bacterial species diversity of spontaneous cocoa bean fermentations carried out at selected farms in Ivory Coast and Brazil. *Food Microbiology*, *28*, 964–973.

Papalexandratou, Z., & De Vuyst, L. (2011b). Assessment of the yeast species composition of cocoa bean fermentations in different cocoa-producing regions using denaturing gradient gel electrophoresis. *FEMS Yeast Research*, *11*, 564–574.

Papalexandratou, Z., Falony, G., Romanens, E., Jimenez, J. C., Amores, F., Daniel, H.-M., & De Vuyst, L. (2011c). Species diversity, community dynamics, and metabolite kinetics of the microbiota associated with traditional Ecuadorian spontaneous cocoa bean fermentations. *Applied and Environmental Microbiology*, *77*, 7698–7714.

Papalexandratou, Z., Lefeber, T., Bahrim, B., Lee, O. S., Daniel, H.-M., & De Vuyst, L. (2013). *Hanseniaspora opuntiae*, *Saccharomyces cerevisiae*, *Lactobacillus fermentum*, and *Acetobacter pasteurianus* predominate during well-performed Malaysian cocoa bean box fermentations, underlining the importance of these microbial species for a successful cocoa . *Food Microbiology*, *35*, 73–85.

Papalexandratou, Z., Vrancken, G., De Bruyne, K., Vandamme, P., & De Vuyst, L. (2011d). Spontaneous organic cocoa bean box fermentations in Brazil are characterized by a restricted species diversity of lactic acid bacteria and acetic acid bacteria. *Food Microbiology*, *28*, 1326–1338.

Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M., & Fire, A. Z. (2007). A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research*, *35*, e130.

Park, E. J., Chun, J., Cha, C. J., Park, W. S., Jeon, C. O., & Bae, J. W. (2012). Bacterial community analysis during fermentation of ten representative kinds of kimchi with barcoded pyrosequencing. *Food Microbiology*, *30*, 197–204.

Peeters, K., Hodgson, D., Convey, P., & Willems, A. (2011). Culturable diversity of heterotrophic bacteria in Forlidas Pond (Pensacola Mountains) and Lundström Lake (Shackleton Range), Antarctica. *Microbial Ecology*, *62*, 399–413.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry*, *55*, 856–866.

Pikuta, E. V., Marsic, D., Bej, A., Tang, J., Krader, P., & Hoover, R. B. (2005). *Carnobacterium pleistocenium* sp. nov., a novel psychrotolerant, facultative anaerobe isolated from permafrost of the Fox Tunnel in Alaska. *International Journal of Systematic and Evolutionary Microbiology*, *55*, 473–478.

Pires, E., Teixeira, J., Brányik, T., & Vicente, A. (2014). Yeast: the soul of beer's aroma-a review of flavour-active esters and higher alcohols produced by the brewing yeast. *Applied Microbiology and Biotechnology*, *98*, 1937–1949.

Pitcher, D. G., Saunders, N. A., & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology*, *8*, 151–156.

Pittet, V., Abegunde, T., Marfleet, T., Haakensen, M., Morrow, K., Jayaprakash, T., Schroeder, K., Trost, B., Byrns, S., Bergsveinson, J., Kusalik, A., & Ziola, B. (2012). Complete genome sequence of the beer spoilage organism *Pediococcus claussenii* ATCC BAA-344^{T}. *Journal of Bacteriology*, *194*, 1271–1272.

Pittet, V., Morrow, K., & Ziola, B. (2011). Ethanol tolerance of lactic acid bacteria, including relevance of the exopolysaccharide gene *gtf*. *Journal of the American Society of Brewing Chemists*, *69*, 57–61.

Pittet, V., Phister, T. G., & Ziola, B. (2013). Transcriptome sequence and plasmid copy number analysis of the brewery isolate *Pediococcus claussenii* ATCC BAA-344^{T} during growth in beer. *PLoS ONE*, *8*, e73627.

Pothakos, V., Snauwaert, C., De Vos, P., Huys, G., & Devlieghere, F. (2014). Psychrotrophic members of *Leuconostoc gasicomitatum*, *Leuconostoc gelidum* and *Lactococcus piscium* dominate at the end of shelf-life in packaged and chilled-stored food products in Belgium. *Food Microbiology*, *39*, 61–67.

Prosser, J. I. (2010). Replicate or lie. *Environmental Microbiology*, *12*, 1806–1810.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*, 7188–7196.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., & Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*, 59–65.

de Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, *102 Suppl*, 6600–6607.

Quince, C., Lanzen, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., Read, L. F., & Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, *6*, 639–641.

Quince, C., Lanzen, A., Davenport, R. J., & Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics*, *12*, 38.

Rademaker, J. L. W., Herbet, H., Starrenburg, M. J. C., Naser, S. M., Gevers, D., Kelly, W. J., Hugenholtz, J., Swings, J., & van Hylckama Vlieg, J. E. T. (2007). Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and $(GTG)_5$-PCR fingerprinting. *Applied and Environmental Microbiology*, *73*, 7128–7137.

Rademaker, J. L. W., Hoolwerf, J. D., Wagendorp, A. A., & te Giffel, M. C. (2006). Assessment of microbial population dynamics during yoghurt and hard cheese fermentation and ripening by DNA population fingerprinting. *International Dairy Journal*, *16*, 457–466.

Rahkila, R., De Bruyne, K., Johansson, P., Vandamme, P., & Björkroth, J. (2014). Reclassification of *Leuconostoc gasicomitatum* as *Leuconostoc gelidum* subsp. *gasicomitatum* comb. nov., description of *Leuconostoc gelidum* subsp. *aenigmaticum* subsp. nov., designation of *Leuconostoc gelidum* subsp. *gelidum* subsp. nov. and emended description. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 1290–1295.

Rappé, M. S., & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Review of Microbiology*, *57*, 369–394.

Raspor, P., & Goranovic, D. (2008). Biotechnological applications of acetic acid bacteria. *Critical Reviews in Biotechnology*, *28*, 101–124.

Ravyts, F., De Vuyst, L., & Leroy, F. (2012). Bacterial diversity and functionalities in food fermentations. *Engineering in Life Sciences*, *12*, 356–367.

Ravyts, F., Vrancken, G., D'Hondt, K., Vasilopoulos, C., De Vuyst, L., & Leroy, F. (2009). Kinetics of growth and 3-methyl-1-butanol production by meat-borne, coagulase-negative staphylococci in view of sausage fermentation. *International Journal of Food Microbiology*, *134*, 89–95.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I., & Saier, M. H. (2012). The major facilitator superfamily (MFS) revisited. *FEBS Journal*, *279*, 2022–2035.

Reeder, J., & Knight, R. (2009). The 'rare biosphere': a reality check. *Nature Methods*, *6*, 636–637.

Richardson, E. J., & Watson, M. (2012). The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*, .

Richter, M., & Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 19126–19131.

Robert, V., Vu, D., Amor, A. B., van de Wiele, N., Brouwer, C., Jabas, B., Szoke, S., Dridi, A., Triki, M., Ben Daoud, S., Chouchen, O., Vaas, L., de Cock, A., Stalpers, J. A., Stalpers, D., Verkley, G. J., Groenewald, M., Dos Santos, F. B., Stegehuis, G., Li, W., Wu, L., Zhang, R., Ma, J., Zhou, M., Gorjon, S. P., Eurwilaichitr, L., Ingsriswang, S., Hansen, K., Schoch, C., Robbertse, B., Irinyi, L., Meyer, W., Cardinali, G., Hawksworth, D. L., Taylor, J. W., & Crous, P. W. (2013). MycoBank gearing up for new horizons. *IMA Fungus*, *4*, 371–379.

Robinson, C. J., Bohannan, B. J., & Young, V. B. (2010). From structure to function: the ecology of host-associated microbial communities. *Microbiology and Molecular Biology Reviews*, *74*, 453–476.

Roh, S. W., Kim, K. H., Nam, Y. D., Chang, H. W., Park, E. J., & Bae, J. W. (2010). Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *ISME Journal*, *4*, 1–16.

Rosselló-Mora, R. (2005). Updating prokaryotic taxonomy. *Journal of Bacteriology*, *187*, 6255–6257.

Rosselló-Mora, R. (2006). DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In E. Stackebrandt (Ed.), *Molecular identification, systematics, and population structure of prokaryotes* chapter 2. (pp. 23–50). Springer, Berlin, Germany.

Rossello-Mora, R., & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiology Reviews*, *25*, 39–67.

Säde, E. (2011). *Leuconostoc spoilage of refrigerated, packaged foods*. Academic dissertation University of Helsinki.

Sakamoto, K., & Konings, W. N. (2003). Beer spoilage bacteria and hop resistance. *International Journal of Food Microbiology*, *89*, 105–124.

Sakamoto, K., Margolles, A., van Veen, H. W., & Konings, W. N. (2001). Hop resistance in the beer spoilage bacterium *Lactobacillus brevis* is mediated by the ATP-binding cassette multidrug transporter HorA. *Journal of Bacteriology*, *183*, 5371–5375.

Sakamoto, N., Tanaka, S., Sonomoto, K., & Nakayama, J. (2011). 16S rRNA pyrosequencing-based investigation of the bacterial community in nukadoko, a pickling bed of fermented rice bran. *International Journal of Food Microbiology*, *144*, 352–359.

Sanchez, J. I., Rossetti, L., Martinez, B., Rodriguez, A., & Giraffa, G. (2006). Application of reverse transcriptase PCR-based T-RFLP to perform semi-quantitative analysis of metabolically active bacteria in dairy fermentations. *Journal of Microbiological Methods*, *65*, 268–277.

Sanger, F., Nicklen, S., & Coulson, A. R. (1992). DNA sequencing with chain-terminating inhibitors. *Biotechnology*, *24*, 104–108.

Santamaria, M., Fosso, B., Consiglio, A., De Caro, G., Grillo, G., Licciulli, F., Liuni, S., Marzano, M., Alonso-Alemany, D., Valiente, G., & Pesole, G. (2012). Reference databases for taxonomic assignment in metagenomics. *Briefings in Bioinformatics*, *13*, 682–695.

Sauer, S., & Kliem, M. (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews Microbiology*, *8*, 74–82.

Sboner, A., Mu, X. J., Greenbaum, D., Auerbach, R. K., & Gerstein, M. B. (2011). The real cost of sequencing: higher than you think! *Genome Biology*, *12*, 125.

Schleifer, K. H. (2009). Classification of bacteria and archaea: Past, present and future. *Systematic and Applied Microbiology*, *32*, 533 – 542.

Schleifer, K. H., & Bergan, T. (1985). Analysis of the chemical composition and primary structure of murein. In *Methods in microbiology* (pp. 123–156). Academic Press volume Volume 18.

Schloss, P. D. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE*, *4*, e8230.

Schloss, P. D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*, *6*, e1000844.

Schloss, P. D., Gevers, D., & Westcott, S. L. (2011a). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE*, *6*, e27310.

Schumann, P. (2011). Peptidoglycan structure. In F. O. A. Rainey (Ed.), *Taxonomy of Prokaryotes* (pp. 101–129). Methods in Microbiology volume 38. (38th ed.).

Seifert, K. A. (2009). Progress towards DNA barcoding of fungi. *Molecular Ecology Resources*, *9*, 83–89.

Seshadri, R., Kravitz, S. A., Smarr, L., Gilna, P., & Frazier, M. (2007). CAMERA: a community resource for metagenomics. *PLoS Biology*, *5*, e75.

Shanta Kumara, H. M. C., & Verachtert, H. (1991). Identification of lambic superattenuating micro-organisms by the use of selective antibiotics. *Journal of the Institute of Brewing*, *97*, 181–185.

Shi, Y., Tyson, G. W., Eppley, J. M., & DeLong, E. F. (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME Journal*, *5*, 999–1013.

Shimwell, J. L. (1937). On the relation between the staining properties of bacteria and their reaction towards hop antiseptic. *Journal of the Institute of Brewing*, *43*, 111–118.

Simon, F., Johan, B., Fredrik, D., Angela, C., Hanlee, J., Katrina, W., & Ronald, W. D. (2007). Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Research*, *35*.

Simpson, W. J. (1993). Ionophoric action of trans-isohumulone on *Lactobacillus brevis*. *Journal of General Microbiology*, *139*, 1041–1045.

Sleator, R. (2011). Phylogenetics. *Archives of Microbiology*, *193*, 235–239.

Smid, W. S., Hugenholtz, J., de Vos, W. M., & J., E. (2006). Safe use of genetically modified lactic acid bacteria in food. Bridging the gap between consumers, green groups, and industry. *Electronic Journal of Biotechnology*, *9*, 1–27.

Smith, M. T., Yamazaki, M., & Poot, G. A. (1990). *Dekkera*, *Brettanomyces* and *Eeniella*: Electrophoretic comparison of enzymes and DNA-DNA homology. *Yeast*, *6*, 299–310.

Snauwaert, I., Hoste, B., De Bruyne, K., Peeters, K., De Vuyst, L., Willems, A., & Vandamme, P. (2013a). *Carnobacterium iners* sp. nov., a psychrophilic, lactic acid-producing bacterium from the littoral zone of an Antarctic pond. *International Journal of Systematic and Evolutionary Microbiology*, *63*, 1370–1375.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 12115–12120.

Spaepen, M., Van Oevelen, D., & Verachtert, H. (1978). Fatty acids and esters produced during the spontaneous fermentation of lambic and gueuze. *Journal of the Institute of Brewing*, *84*, 278–282.

Spaepen, M., Van Oevelen, D., & Verachtert, H. (1979). Higher fatty acid (HFA) and HFA-ester content of spontaneously fermented Belgian beers and evaluation of their analytical determination. *Brauwissenschaft*, *32*, S1–S6.

Spaepen, M., & Verachtert, H. (1982). Esterase activity in the genus *Brettanomyces*. *Journal of the Institute of Brewing*, *88*, 11–17.

Spitaels, F., Li, L., Wieme, A., Balzarini, T., Cleenwerck, I., Van Landschoot, A., De Vuyst, L., & Vandamme, P. (2014a). *Acetobacter lambici* sp. nov., isolated from fermenting lambic beer. *International Journal of Systematic and Evolutionary Microbiology*, *64*, 1083–1089.

Spitaels, F., Wieme, A., Janssens, M., Aerts, M., Daniel, H.-M., Van Landschoot, A., De Vuyst, L., & Vandamme, P. (2014b). The microbial diversity of traditional spontaneously fermented lambic beer. *PLoS ONE*, *9*, e95384.

Stackebrandt, E., & Ebers, J. (2006). Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today*, *33*, 152.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., KÃďmpfer, P., Maiden, M. C. J., Nesme, X., RossellÃş-Mora, R., Swings, J., TrÃijper, H. G., Vauterin, L., Ward, A. C., & Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *52*, 1043–7.

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic note: a place for dna-dna reassociation and 16s rrna sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*, *44*, 846–849.

Staley, J. T. (2006). The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 1899–1909.

Stiles, M. E., & Holzapfel, W. H. (1997). Lactic acid bacteria of foods and their current taxonomy. *International Journal of Food Microbiology*, *36*, 1–29.

Suárez, R., Suárez-Lepe, J., Morata, A., & Calderón, F. (2007). The production of ethylphenols in wine by yeasts of the genera *Brettanomyces* and *Dekkera*: A review. *Food Chemistry*, *102*, 10–21.

Suenaga, H. (2012). Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environmental Microbiology*, *14*, 13–22.

Sugawara, H., Abe, T., Gojobori, T., & Tateno, Y. (2007). DDBJ working on evaluation and classification of bacterial genes in INSDC. *Nucleic Acids Research*, *35*, D13–5.

Sutcliffe, I. C., Trujillo, M. E., & Goodfellow, M. (2012). A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie van Leeuwenhoek*, *101*, 13–20.

Suzuki, K., Ozaki, K., & Yamashita, H. (2004). Comparative analysis of conserved genetic markers and adjacent DNA regions identified in beer-spoilage lactic acid bacteria. *Letters in Applied Microbiology*, *39*, 240–245.

Sybesma, W., Hugenholtz, J., de Vos, W. M., & Smid, E. J. (2006). Safe use of genetically modified lactic acid bacteria in food: Bridging the gap between consumers, green groups, and industry. *Electronic Journal of Biotechnology*, *9*.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and g+c-content biases. *Molecular Biology and Evolution*, *9*, 678–687.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*, 2731–2739.

Tanasupawat, S., Shida, O., Okada, S., & Komagata, K. (2000). *Lactobacillus acidipiscis* sp. nov. and *Weissella thailandensis* sp. nov., isolated from fermented fish in Thailand. *International Journal of Systematic and Evolutionary Microbiology*, *50*, 1479–1485.

Teeling, H., Fuchs, B. M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C. M., Kassabgy, M., Huang, S., Mann, A. J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F. D., Callies, U., Gerdts, G., Wichels, A., Wiltshire, K. H., Glöckner, F. O., Schweder, T., & Amann, R. (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*, *336*, 608–611.

Teeling, H., & Glöckner, F. O. (2012). Current opportunities and challenges in microbial metagenome analysisâĂŤa bioinformatic perspective. *Briefings in Bioinformatics*, *13*, 728–742.

Temperton, B., & Giovannoni, S. J. (2012). Metagenomics: microbial diversity through a scratched lens. *Current Opinion in Microbiology*, *15*, 605 – 612.

Teuber, M., & Schmalreck, A. F. (1973). Membrane leakage in *Bacillus subtilis* 168 induced by the hop constituents lupulone, humulone, isohumulone and humulinic acid. *Archiv für Mikrobiologie*, *94*, 159–171.

Thompson, J. R., Marcelino, L. A., & Polz, M. F. (2002). Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Research*, *30*, 2083–2088.

Tilden, E. B., & Svec, M. (1952). Further studies of a differential culture technique for estimations of acidogenic bacteria in saliva. II. Species of lactobacilli isolated from saliva and their distribution in a group of children. *Journal of Dental Research*, *31*, 831–838.

Timke, M., Wang-Lieu, N. Q., Altendorf, K., & Lipski, A. (2008). Identity, beer spoiling and biofilm forming potential of yeasts from beer bottling plant associated biofilms. *Antonie van Leeuwenhoek*, *93*, 151–161.

Tindall, B. J., Rossello-Mora, R., Busse, H. J., Ludwig, W., & Kampfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology*, *60*, 249–266.

Tiukova, I. A., Petterson, M. E., Tellgren-Roth, C., Bunikis, I., Eberhard, T., Pettersson, O. V., & Passoth, V. (2013). Transcriptome of the alternative ethanol production strain *Dekkera bruxellensis* CBS 11270 in sugar limited, low oxygen cultivation. *PLoS ONE*, *8*, e58455.

Torriani, S., Gatto, V., Sembeni, S., Tofalo, R., Suzzi, G., Belletti, N., Gardini, F., & Bover-Cid, S. (2008). Rapid detection and quantification of tyrosine decarboxylase gene (*tdc*) and its expression in gram-positive bacteria associated with fermented foods using PCR-based methods. *Journal of Food Protection*, *71*, 93–101.

Turenne, C. Y., Tschetter, L., Wolfe, J., & Kabani, A. (2001). Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous *Mycobacterium* species. *Journal of Clinical Microbiology*, *39*, 3637–3648.

Turnbaugh, P. J., & Gordon, J. I. (2008). An invitation to the marriage of metagenomics and metabolomics. *Cell*, *134*, 708–713.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., & Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, *428*, 37–43.

van der Aa Kuhle, A., & Jespersen, L. (1998). Detection and identification of wild yeasts in lager breweries. *International Journal of Food Microbiology*, *43*, 205–213.

Van Oevelen, D., L'Escaille, F., & Verachtert, H. (1976). Synthesis of aroma components during the spontaneous fermentation of lambic and gueuze. *Journal of the Institute of Brewing*, *82*, 322–326.

Van Oevelen, D., Spaepen, M., Timmermans, P., & Verachtert, H. (1977). Microbiological aspects of spontaneous wort fermentation in the production of lambic and gueuze. *Journal of the Institute of Brewing*, *83*, 356–360.

Van Oevelen, D., & Verachtert, H. (1979). Slime production by brewery strains of *Pediococcus cerevisiae*. *Journal of the American Society of Brewing Chemists*, *37*, 34–37.

Vancanneyt, M., Mengaud, J., Cleenwerck, I., Vanhonacker, K., Hoste, B., Dawyndt, P., Degivry, M. C., Ringuet, D., Janssens, D., & Swings, J. (2004). Reclassification of *Lactobacillus kefirgranum* Takizawa et al. 1994 as *Lactobacillus kefiranofaciens* subsp. *kefirgranum* subsp. nov. and emended description of *L. kefiranofaciens* Fujisawa et al. 1988. *International Journal of Systematic and Evolutionary Microbiology*, *54*, 551–556.

Vandamme, P., & Peeters, C. (2014). Time to revisit polyphasic taxonomy. *Antonie Van Leeuwenhoek*, .

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews*, *60*, 407–438.

Varley, K. E., & Mitra, R. D. (2008). Nested patch PCR enables highly multiplexed mutation discovery in candidate genes. *Genome Research*, *18*, 1844–1850.

Vela, A. I., Fernández, A., Bernaldo de Quirós, Y., Herráez, P., Domínguez, L., & Fernández-Garayzábal, J. F. (2011). *Weissella ceti* sp. nov., isolated from beaked whales (*Mesoplodon bidens*). *International Journal of Systematic and Evolutionary Microbiology*, *61*, 2758–2762.

Verachtert, H., & Derdelinckx, G. (2005). Acidic beers: enjoyable reminiscences of the past. *Cerevisia, Belgian Journal of Brewing and Biotechnology*, *30*, 38–47.

Verachtert, H., & Iserentant, D. (1995). Properties of Belgian acid beers and their microflora. Part I. The production of gueuze and related refreshing acid beers. *Cerevisia, Belgian Journal of Brewing and Biotechnology*, *20*, 37–41.

Verstrepen, K. J., Derdelinckx, G., Dufour, J. P., Winderickx, J., Thevelein, J. M., Pretorius, I. S., & Delvaux, F. R. (2003). Flavor-active esters: adding fruitiness to beer. *Journal of Bioscience and Bioengineering*, *96*, 110–118.

Vieites, J. M., Guazzaroni, M. E., Beloqui, A., Golyshin, P. N., & Ferrer, M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews*, *33*, 236–255.

Vihavainen, E. J., Murros, A. E., & Björkroth, K. J. (2008). *Leuconostoc* spoilage of vacuum-packaged vegetable sausages. *Journal of Food Protection*, *71*, 2312–2315.

Walker, A. W., Duncan, S. H., Louis, P., & Flint, H. J. (2014). Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends in Microbiology*, *22*, 267–274.

Walling, E., Gindreau, E., & Lonvaud-Funel, A. (2005). A putative glucan synthase gene *dps* detected in exopolysaccharide-producing *Pediococcus damnosus* and *Oenococcus oeni* strains isolated from wine and cider. *International Journal of Food Microbiology*, *98*, 53–62.

Wang, D. G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., & Lander, E. S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, *280*, 1077–1082.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve bayesian classifier for rapid assignment of rRNA Sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*, 5261–5267.

Wang, X., Eijkemans, M. J. C., Wallinga, J., Biesbroek, G., TrzciÅĎski, K., Sanders, E. A. M., & Bogaert, D. (2012). Multivariate approach for studying interactions between environmental variables and microbial communities. *PLoS ONE*, *7*, e50267.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., Gillespie, J. J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E. K., Olson, R., Overbeek, R., Pusch, G. D., Shukla, M., Schulman, J., Stevens, R. L., Sullivan, D. E., Vonstein, V., Warren, A., Will, R., Wilson, M. J. C., Yoo, H. S., Zhang, C., Zhang, Y., & Sobral, B. W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*, *42*, D581–D591.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P., & Truper, H. G. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology*, *37*, 463–464.

Weiss, S., Samson, F., Navarro, D., & Casaregola, S. (2013). YeastIP: a database for identification and phylogeny of Saccharomycotina yeasts. *FEMS Yeast Research*, *13*, 117–125.

White, T., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. Innis, D. Gelfand, J. Shinsky, & T. White (Eds.), *PCR protocols: a guide to methods and applications* (pp. 315–322). Academic Press.

Wieme, A., Cleenwerck, I., Van Landschoot, A., & Vandamme, P. (2012). *Pediococcus lolii* DSM 19927^{T} and JCM 15055^{T} are strains of *Pediococcus acidilactici*. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 3105–3108.

Wieme, A., Spitaels, F., Aerts, M., De Bruyne, K., Van Landschoot, A., & Vandamme, P. (2014). Effects of growth medium on matrix-assisted laser desorption-ionization time of flight mass spectra: a case study of acetic acid bacteria. *Applied and Environmental Microbiology*, *80*, 1528–1538.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L., & Koonin, E. V. (2001). Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evolutionary Biology*, *1*, 8.

Xufre, A., Albergaria, H., Inacio, J., Spencer-Martins, I., & Girio, F. (2006). Application of fluorescence *in situ* hybridisation (FISH) to the analysis of yeast population dynamics in winery and laboratory grape must fermentations. *International Journal of Food Microbiology*, *108*, 376–384.

Yasuhara, T., Yuuki, T., & Kagami, N. (2001). Novel quantitative method for detection of *Pectinatus* using rRNA targeted fluorescent probes. *Journal of the American Society of Brewing Chemists*, *59*, 121–177.

Yilmaz, P., Gilbert, J. A., Knight, R., Amaral-Zettler, L., Karsch-Mizrachi, I., Cochrane, G., Nakamura, Y., Sansone, S. A., Glockner, F. O., & Field, D. (2011). The genomic standards consortium: bringing standards to life for microbial ecology. *ISME Journal*, *5*, 1565–1567.

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., & Glöckner, F. O. (2013). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, .

Yoon, B., Jang, S., & Chang, H.-I. (2011). Sequence analysis of the *Lactobacillus* temperate phage Sha1. *Archives of Virology*, *156*, 1681–1684.

You, Z., Yongjie, L., Karlene, H. L., Jonathan, J. D., & David, S. W. (2011). PHAST: a fast phage search tool. *Nucleic Acids Research*, *39*, W347–W352.

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*, 847–848.

Ze, X., Le Mougen, F., Duncan, S. H., Louis, P., & Flint, H. J. (2013). Some are more equal than others: the role of 'keystone' species in the degradation of recalcitrant substrates. *Gut microbes*, *4*, 236–40.

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*, 821–829.

Zhang, Z. G., Ye, Z. Q., Yu, L., & Shi, P. (2011). Phylogenomic reconstruction of lactic acid bacteria: an update. *BMC Evolutionary Biology*, *11*.

Part VI

Curriculum Vitae

Isabel Snauwaert

Gulden-Peerdenstraat 45 • 8310 Assebroek • Belgium
Phone: +32 (0) 471 87 12 54 • E-mail: isabelsnauwaert@hotmail.com
Nationality: Belgian • Date of birth: the 29th of April 1987
Married with Bram Gilté

Objective

I am an **ambitious** and **result-oriented team player** with a background in Biotechnology and experience in Genomics, Microbiology, and Bioinformatics. My objective is to work in an **international** and **challenging scientific environment** where my data handling, organizational, and communication skills can be effectively employed.

Experience

I am currently finishing my PhD project under the supervision of Prof. Dr. Peter Vandamme at the Laboratory of Microbiology, Ghent University and Prof. dr. ir. Luc De Vuyst at the Research Group of Industrial Microbiology and Food Biotechnology, Vrije Universiteit Brussel.

I carried out research on different aspects of microbiology, including microbial diversity analysis, comparative genomics, metabolite target analysis, taxonomy, and microbial ecology. I am familiar with the following methods: next-generation sequencing technologies and data handling, comparative genomics, mass spectrometry, chromatography, DNA extraction, cultivation, phylogenetic analysis using different software packages, PCR-based methods, and other. Details on my experience can be found below.

Education

Master of Biochemistry and Biotechnology (*magma cum laude*), Ghent University

- Specialization: Microbial biotechnology and biomedical biotechnology

Bachelor of Biochemistry and Biotechnology (*cum laude*), Ghent University
Secondary school degree, Sint-Andreaslyceum Sint-Kruis, Bruges

- Specialization in science and mathematics

Skills

Language skills

- Dutch: mother tongue; English: proficient user; French: basic user; Portuguese: basic user

Social and organizational skills

- Team work through experience in various disciplines including the youth movement (Chiro: member for 14 years of which 6 years as a leader), scientific popularization (‘Dag van de wetenschap’: 2012 and 2013), and other.

Computer skills

- Excellent command of Microsoft Windows and Office (Excel, Word, PowerPoint and Outlook); mothur software package (command line interface); Internet applications
- Command of R software package
- Basics of Unix, Latex, High-Performance computing, and Python

Other skills

- Sports: triathlon (Olympic distance), running, bike racing, swimming, 4 years safeguard at the beach, 2 years safeguard in an indoor swimming pool; hobby brewer

Experience: details

September 2010- now

Analysis of the microbial diversity and metabolite composition of Belgian acidic ales

- Knowledge on DNA extraction, PCR optimizations, 454 pyrosequencing, data analysis using the mothur software package, and statistical analysis using the R software package. Contributed to the first use of this technology and its resulting data analysis in the laboratory.
- Collaboration with the laboratory of Prof. Dr. Ir. Luc De Vuyst (Research Group of Industrial Microbiology and Food Biotechnology, Department of Bioengineering Sciences, Vrije Universiteit Brussel) and Prof. Dr. Anita Van Landschoot (Laboratory of Biochemistry and Brewing, Faculty of Bioscience Engineering, Ghent University)
- Example publication (not yet submitted)
    - Snauwaert I, Roels S, Van Nieuwerburg F, Van Landschoot A, De Vuyst L, Vandamme P (2014) Microbial diversity and metabolite composition of Belgian red-brown acidic ales. In preparation.

Comparative genome analysis of a *Pediococcus damnosus* strain isolated from beer

- Knowledge on Illumina sequencing, data analysis using different open source resources (RAST, IMG-RE, PGAAP, MAUVE, ACT) and CLC genomics workbench, interpretation of the annotations found in the genome sequences, and other.
- Example publication (net yet submitted)
    - Snauwaert I, Stragier P, De Vuyst L, Vandamme P (2014) Comparative genome analysis of *Pediococcus damnosus* LMG 28219, a strain well-adapted to the brewery environment. In preparation.

Taxonomic study

- Knowledge on mass spectrometry, multi locus sequence analysis, phylogeny, general cultivation, and other. Contributed to the setting up of a mass spectral database containing protein profiles of lactic acid bacteria.
- Example publications
    - Snauwaert I, Hoste B, De Bruyne K, Peeters K, De Vuyst L, et al. (2013) *Carnobacterium iners* sp. nov., a psychrophilic, lactic acid-producing bacterium from the littoral zone of an Antarctic pond. International Journal of Systematic and Evolutionary Microbiology 63: 1370-1375.
    - Snauwaert I, Papalexandratou Z, De Vuyst L, Vandamme P (2013) Characterization of strains of *Weissella fabalis* sp. nov. and *Fructobacillus tropaeoli* from spontaneous cocoa bean fermentations. International Journal of Systematic and Evolutionary Microbiology 63: 1709-1716.

Collaboration to other projects

- Spitaels F, Van Kerrebroeck S, Wieme A, Snauwaert I, Aerts M, Van Landschoot A, De Vuyst L, Vandamme P (2014). Microbiota and metabolites of aged bottled gueuze beers converge to the same composition. In preparation.
- Lyhs U, Snauwaert I, Pihlajaviita S, De Vuyst L, Vandamme P (2014) *Leuconostoc rapi* sp. nov., isolated from sous-vide cooked rutabaga. In preparation.
- Huch M, De Bruyne K, Cleenwerck I, Bub A, Cho G, Watzl B, Snauwaert I, Franz C, Vandamme P (2013) *Streptococcus rubneri* sp. nov., isolated from the human throat. International Journal of Systematic and Evolutionary Microbiology 63:4026-4032
- Nguyen D, Cnockaert M, Van Hoorde K, De Brandt E, Snauwaert I, Snauwaert C, De Vuyst L, Thanh Le B, Vandamme P (2012) *Lactobacillus porcinae* sp. nov. isolated from traditional Vietnamese nem chua. International Journal of Systematic and Evolutionary Microbiology 0.044123-0

Guidance of students

- Mieke Vandewalle and Annelies Torfs (2nd Master in Biochemistry and Biotechnology); Elisa Dobbelaere (Master Pharmaceutical sciences); Serge Tambue (1st Master Biochemistry and Biotechnology); several practical courses.

Workshops followed in the scope of the PhD

- Mothur workshop 2012 and R workshop 2013, Detroit area (Michigan, USA) organized by Dr. Patrick Schloss; COST action workshops on 'Microbial ecology & the earth system: collaboration for insight and success with the new generation of sequencing tools' in Liverpool and 'Workshop on Linking microbial ecology & evolution' in Berlin.

Courses followed in the scope of the PhD (doctoral program Ghent University)

- Getting started with high-performance computing (included basics of Unix and Python); Introduction to Statistics; Brewing technology
- Advanced academic English: writing skills; Communication skills: Conflict handling and Meeting skills; Project management; Academic posters

Posters on symposia

- Mass spectrometry in Food and Feed (9 June 2011), Merelbeke, Belgium
- 10th symposium on lactic acid bacteria (28 August -1 September 2011), Egmond aan zee, The Netherlands
- ISME 14 congress (21-24 of august 2012), Copenhagen, Denmark

January 2010- June 2010

2nd Master thesis

- 'Gut microbiota of Cystic fibrosis patients and their healthy siblings' under the supervision of Prof. Dr. Peter Vandamme (Laboratory of Microbiology, Ghent University)